

Direction de la recherche, des études,  
de l'évaluation et des statistiques  
DREES

SÉRIE  
SOURCES ET MÉTHODES

**DOCUMENT  
DE  
TRAVAIL**

Méthologie des projections  
d'effectifs de médecins

Ketty Attal-Toubert - Mélanie Vanderschelden

n° 7 – novembre 2009

## Résumé

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.

Une nouvelle version du modèle a permis d'établir les résultats du scénario tendanciel à l'été 2008, et a été utilisée pour réaliser un nouvel exercice de projections en 2008-2009. Les premiers résultats de cet exercice de projections ont été publiés dans le n° 679 d'Études et Résultats (février 2009) et une présentation de résultats plus détaillés figure dans le n° 12 des Dossiers Solidarité Santé (novembre 2009).

Nous présentons dans ce document de travail la version 2008 du modèle de projections d'effectifs de médecins, c'est-à-dire les méthodes et les sources de données utilisées, la structure du modèle, les choix effectués, les hypothèses faites et les résultats produits.

En complément à ce document de travail, on pourra lire dans les actes des journées de méthodologie statistique de 2009 la communication intitulée : « Mise en œuvre de la microsimulation : l'exemple du modèle de projections d'effectifs de médecins », accessible en ligne.

## Sommaire

<b>Résumé.....</b>	<b>3</b>
<b>A – Présentation générale du modèle .....</b>	<b>7</b>
1. La logique de construction du modèle.....	7
a. Les objectifs du modèle.....	7
b. Le champ retenu.....	9
c. Les événements modélisés .....	10
d. Les méthodes : modélisation agrégée et microsimulation.....	12
2. Le choix des sources .....	15
a. Les sources de données sur les médecins actifs : ADELI et les fichiers de l'Ordre.....	15
b. Les sources de données sur les étudiants en médecine.....	18
3. Le choix des hypothèses.....	22
4. L'enchaînement des calculs.....	24
5. Les résultats produits par le modèle .....	29
<b>B – Module 1 - Les études médicales .....</b>	<b>30</b>
1. Sous-module 1.1 : Du PCEM 2 (deuxième et dernière année de premier cycle des études médicales) au DCEM 4 (quatrième et dernière année de deuxième cycle).....	30
a. Le choix de la méthode .....	30
b. La simulation : du numerus clausus aux effectifs en DCEM 4 .....	32
c. À propos des taux.....	35
d. Les simulations possibles.....	37
2. Sous-module 1.2 : du DCEM 4 au TCEM 1 (première année de troisième cycle des études médicales) ou l'affectation des étudiants en troisième cycle à l'issue des ECN .....	38
a. Le choix de la méthode .....	38
b. La simulation : des candidats aux ECN aux affectations en TCEM 1 .....	38
c. À propos des taux et des répartitions estimées .....	41
d. Les simulations possibles.....	43
3. Sous-module 1.3 : du TCEM 1 (première année de troisième cycle des études médicales) à l'obtention du diplôme.....	44
a. Le choix de la méthode .....	44
b. La simulation : des affectations en TCEM 1 aux diplômés.....	44
c. À propos des taux et des répartitions estimées .....	47
d. Les simulations possibles.....	49
<b>C – Module 2 – L'entrée dans la vie active .....</b>	<b>49</b>
1. Sous-module 2.1 : le calcul des effectifs de médecins qui exerceront la médecine et l'attribution d'un délai à l'entrée dans la vie active .....	50
a. Le « taux de fuites » ou le taux de médecins qui n'exerceront jamais la médecine .....	50
b. L'attribution d'un délai à l'entrée dans la vie active.....	50
c. Le cas particulier des médecins diplômés au cours de l'une des cinq années précédant le début de la période de projection .....	55
d. Création de la base des médecins à faire entrer dans la vie active au cours de l'année .....	56
2. Sous-module 2.2 : attribution d'un sexe et d'un âge aux médecins entrant dans la vie active au cours de l'année .....	57
3. Sous-module 2.3 : attribution d'une région d'exercice aux médecins entrant dans la vie active au cours de l'année .....	58
4. Sous-module 2.4 : attribution d'un mode d'exercice aux médecins entrant dans la vie active au cours de l'année.....	60
5. Sous-module 2.5 : attribution d'une zone d'exercice aux médecins entrant dans la vie active au cours de l'année.....	61
6. Les simulations possibles pour le module 2 .....	62

<b>D – Module 3 – Les changements au cours de la vie active .....</b>	<b>63</b>
1. Sous-module 3.1 : le changement de région d'exercice .....	63
a. Le choix de la méthode .....	63
b. L'estimation des probabilités et la simulation des changements de région.....	63
2. Sous-module 3.2 : le changement de mode d'exercice.....	65
3. Sous-module 3.3 : le changement de zone d'exercice .....	66
a. Le médecin reste en métropole.....	66
b. Le médecin quitte un DOM, un TOM ou la Corse, pour la métropole.....	67
4. Sous-module 3.4 : la cessation temporaire d'activité .....	68
5. Sous-module 3.5 : la reprise d'activité .....	69
a. La sélection des médecins reprenant une activité.....	69
b. L'attribution d'un mode, d'une région et d'une zone d'exercice aux médecins reprenant une activité .....	70
c. Fin de la simulation des reprises d'activité .....	71
6. Les simulations possibles pour le module 3 .....	71
<b>E – Module 4 – La sortie définitive de la vie active .....</b>	<b>71</b>
1. Sous-module 4.1 : le départ à la retraite .....	71
2. Sous-module 4.2 : le décès .....	73
3. Les simulations possibles pour le module 4 .....	73

## A – Présentation générale du modèle

### 1. La logique de construction du modèle

#### a. Les objectifs du modèle

❶ Le modèle produit des **effectifs projetés de médecins actifs pour chaque année de 2007 à 2030 par sexe, âge, spécialité, région d'exercice, mode d'exercice et zone d'exercice.**

Les effectifs de certaines spécialités sont très faibles. Par conséquent, ils ont été regroupés. Les effectifs projetés sont donc déclinés selon les **23 spécialités** suivantes :

- anesthésie-réanimation
- biologie médicale
- cardiologie
- chirurgie
- anatomie et cytologie pathologiques
- dermatologie-vénérologie
- endocrinologie
- gastro-entérologie
- gynécologie-obstétrique
- médecine générale
- médecine du travail
- médecine interne
- neurologie
- ophtalmologie
- ORL
- pédiatrie
- pneumologie
- psychiatrie
- radiologie
- rééducation et réadaptation fonctionnelle
- rhumatologie
- santé publique
- autres spécialités

Pour les exercices de projections précédents, les effectifs de spécialistes en **médecine du travail** et en **santé publique** étaient regroupés. Or, premièrement, les effectifs correspondant à chacune de ces spécialités sont assez conséquents ; deuxièmement, ces deux spécialités sont très différentes ; troisièmement, aux ECN (épreuves classantes nationales), ces deux spécialités constituent deux disciplines distinctes. La version 2008 du modèle permet donc de projeter séparément les effectifs correspondant à la médecine du travail et à la santé publique.

En revanche, bien que la **gynécologie médicale** constitue une discipline des ECN, les effectifs de cette spécialité ont été regroupés avec ceux des « autres spécialités », en raison de leur petite taille.

Pour le **mode d'exercice**, les modalités suivantes ont été retenues :

- libéral
- salarié hospitalier
- salarié non hospitalier
- mixte

La modalité « mixte » correspond aux médecins qui ne sont ni libéraux exclusifs, ni salariés hospitaliers exclusifs, ni salariés non hospitaliers exclusifs. Elle ne correspond donc pas uniquement aux cas des médecins combinant une activité libérale et une activité salariée : un médecin à la fois salarié hospitalier et salarié non hospitalier a également un mode d'exercice « mixte » (au sens du modèle).

Il n'a pas été possible de retenir la notion de mode d'exercice principal (correspondant à l'activité principale d'ADELI), la source retenue (voir A. 2.) ne permettant pas de déterminer l'activité principale. Décomposer la modalité « mixte » en plusieurs modalités pour rendre les résultats plus « parlants » aurait conduit à projeter des effectifs trop faibles.

Pour la **zone d'exercice**, ont été retenues les modalités suivantes :

- pôle urbain avec CHU
- pôle urbain sans CHU
- commune monopolarisée (couronne périurbaine) ou multipolarisée
- espace à dominante rurale

Pour les exercices précédents de projections, les effectifs projetés de médecins étaient répartis non pas selon le type de zone d'exercice, mais selon la taille d'unité urbaine de leur commune d'exercice. Cette variable comptait seulement trois modalités (moins de 5 000 habitants, de 5 000 à 99 999 habitants, 100 000 habitants ou plus), pour que les effectifs de médecins correspondant à chaque modalité soient suffisamment conséquents. Cependant, une part importante des médecins exerçaient dans une unité urbaine de taille intermédiaire. Un nouveau zonage a donc été construit à partir du zonage en aires urbaines de l'Insee qui mesure l'influence urbaine à partir des navettes domicile-travail. Il a été retenu pour son utilisation aisée (la correspondance entre le code commune Insee et le type de zone est établi par l'Insee) et l'intérêt qu'il présente pour l'analyse des résultats des projections

La distinction entre pôle urbain avec CHU et pôle urbain sans CHU a été faite à la demande d'utilisateurs des projections, qui s'intéressaient à la proximité entre le lieu d'exercice du médecin et les CHU environnants. L'utilisation du zonage en aires urbaines et la distinction entre pôles urbains avec ou sans CHU permet de rendre compte assez nettement de la proximité entre le lieu d'exercice du médecin et le CHU, tout en évitant le calcul d'une distance entre le lieu d'exercice du médecin et le CHU le plus proche, calcul possible, mais un peu complexe et assez coûteux (en temps de calcul).

Il aurait été possible de distinguer les médecins exerçant dans une commune monopolarisée des médecins exerçant dans une commune multipolarisée. Cependant, les effectifs correspondant auraient été trop faibles.

Les effectifs projetés ne sont cependant pas déclinés par zone d'exercice pour les DOM, les TOM et la Corse. Le zonage en aires urbaines n'est pas défini pour les DOM-TOM. Il existe en revanche un zonage en aires urbaines pour la Corse, mais les effectifs de médecins de la région sont faibles.

Afin de parvenir au découpage présenté ci-dessus, le code Insee de la commune d'exercice du médecin est associé au code de l'aire urbaine et au type d'aire urbaine correspondants (commune multipolarisée, espace à dominante rurale, etc.), grâce à une table de correspondance. Pour les médecins dont le code de la commune d'exercice correspond à un « pôle urbain », la distinction entre pôle urbain avec CHU et pôle urbain sans CHU est faite à partir du code d'aire urbaine : on attribue aux médecins pour lesquels la commune d'exercice fait partie d'un pôle urbain contenant un CHU la modalité « pôle urbain avec CHU », et aux autres médecins la modalité « pôle urbain sans CHU ».

② Le modèle de projections produit également **des informations permettant de comprendre les évolutions des effectifs projetés** : par exemple, il donne des informations sur les entrées et sorties de la vie active au cours de la période de projection.

③ Il prend aussi en compte la **diversité des comportements** des médecins en fonction de leurs caractéristiques, par exemple des médecins d'âge ou de spécialité différents n'auront pas la même probabilité de changer de mode d'exercice.

④ Enfin, la simulation de **différents scénarios** est possible, par exemple on peut faire varier le niveau du *numerus clausus*.

### ***b. Le champ retenu***

Le champ retenu pour les projections est l'ensemble des **médecins en activité en France** (métropole + DOM-TOM), qu'ils aient une activité régulière ou qu'ils soient remplaçants. Les **retraités ayant conservé ou repris une petite activité** sont **exclus** du champ. Sont également exclus du champ les médecins en **cessation temporaire d'activité**. Enfin, ont été ôtés des effectifs les **médecins âgés de plus de 80 ans** (strictement).

Pour les exercices de projections précédents, le champ retenu était celui des médecins en activité en France métropolitaine et dans les DOM, y compris les remplaçants, mais aussi les médecins en cessation temporaire d'activité, que la source utilisée (ADELI) ne permettait pas de repérer, et donc d'ôter des effectifs projetés. Les médecins exerçant dans les **TOM** n'étaient pas pris en compte, ADELI ne les répertoriait pas. La source utilisée en 2008 (voir A. 2.) permet de distinguer les médecins en cessation temporaire d'activité et répertorie les médecins exerçant dans les TOM. En revanche, les retraités et les médecins de plus de 80 ans étaient déjà exclus du champ pour les exercices de projections précédents. Il s'agit en effet de projeter des effectifs de médecins exerçant de manière régulière la médecine. Les médecins de plus de 80 ans sont rares dans les fichiers pouvant servir de sources.

Jusqu'en 2004, les 4 **DOM** constituaient une seule « région » pour les projections d'effectifs de médecins, et les effectifs de médecins exerçant en **Corse** étaient regroupés avec les effectifs de médecins exerçant dans la région PACA. Ces regroupements étaient motivés par les faibles effectifs correspondant à chaque DOM et à la Corse, mais l'analyse des résultats était difficile. La version 2008 du modèle de projections distingue les Antilles-Guyane de la Réunion, et la Corse de la région PACA. Les TOM (regroupés) constituent également une région. Dans la réalité, la Guadeloupe, la Martinique et la Guyane constituent trois régions distinctes, mais les effectifs correspondant à chacun de ces trois DOM sont trop faibles pour être projetés séparément.

Il faut toutefois rester prudent sur l'interprétation des résultats pour les DOM et la Corse, les effectifs restant assez faibles.

Enfin, il n'est pas tenu compte pour les projections (sauf pour certaines variantes) des **migrations internationales de médecins au cours de leur période d'activité**, qu'il s'agisse des flux de médecins entrants (médecins diplômés à l'étranger venant exercer en France) ou des flux de médecins sortants (médecins diplômés en France partant exercer à l'étranger). Pourtant, certains médecins diplômés à l'étranger sont autorisés à exercer en France. Ce choix, qui a été longuement débattu, est motivé par deux raisons :

- la très forte incertitude quant aux flux de médecins futurs (notamment pour les flux entrants), ces flux étant dépendants des politiques conduites en la matière. Le choix des hypothèses qui pourraient être faites sur ces flux sur le long terme est délicat.

- la difficulté à trouver les données nécessaires à une modélisation correcte de ces flux. Les données qui permettraient de modéliser correctement les flux entrants font en partie défaut. Il faut connaître le sexe, l'âge, la spécialité, la région d'exercice, le mode d'exercice, la zone d'exercice des médecins qui viennent exercer en France, dont on sait que le profil est particulier, et variable selon les politiques conduites. Par exemple, certaines spécialités peuvent être momentanément ciblées. Or, les informations sur les flux actuels sont incomplètes. De plus, le choix des hypothèses qu'il faudrait faire sur les caractéristiques des médecins qui viendraient à l'avenir exercer en France est lui aussi très délicat. Quant aux flux sortants, aucune donnée fiable ne permet de les modéliser (la liste spéciale du CNOM est insuffisante).

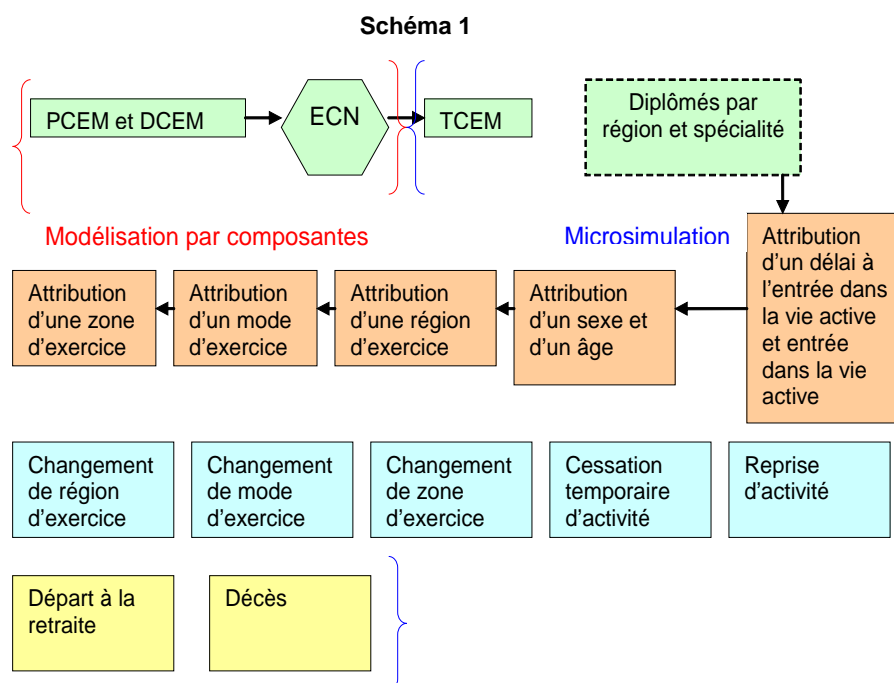
En revanche, les arrivées de l'étranger et les départs d'étudiants inscrits dans les facultés de médecine françaises vers l'étranger sont pris en compte, via les taux d'arrivées et les taux d'abandons en cours d'études (voir A. 2. b. et B. 2. c.). De même, les départs vers l'étranger, avant leur entrée dans la vie active, de médecins diplômés en France sont pris en compte, par l'intermédiaire du taux de fuites (voir C. 1. a.). Les médecins diplômés à l'étranger et exerçant en France au début de la période de projection sont comptabilisés dans les effectifs projetés dès lors qu'ils étaient inscrits à l'Ordre au 31/12/2006 (pour l'exercice de projections 2008).

### c. Les événements modélisés

Le choix des événements modélisés a été dicté par le degré de précision souhaité, mais aussi et surtout par les attentes quant aux possibilités du modèle en matière de simulation.

Ainsi, dans la version précédente du modèle, la mobilité géographique des médecins en cours de carrière n'était pas modélisée. Or, les attentes à l'égard du modèle en ce qui concerne la précision des effectifs projetés au niveau régional justifiaient que l'on tente de modéliser la mobilité géographique des médecins si son ampleur s'avérait conséquente. La modélisation du déroulement des études médicales s'imposait quant à elle du fait du besoin clairement identifié de simuler les effets de politiques de régulation portant sur le niveau du *numerus clausus* national ou sur la répartition des postes ouverts aux ECN, par exemple.

Finalement, ont été modélisés les événements figurant sur le schéma suivant :





Le **déroulement des études médicales** a donc 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 ECN. La modélisation des ECN, dont le fonctionnement est assez complexe, a été réalisée (même si la modélisation retenue présente des limites - voir B. 3.), parce qu'il devait être possible de simuler entre autres les effets de modifications de la répartition des postes ouverts aux ECN (par exemple une modification de la part des postes ouverts en médecine générale dans l'ensemble des postes ouverts). En revanche, la modélisation du déroulement des études médicales ne commence pas au moment de l'entrée en première année de premier cycle, mais au moment de l'entrée en deuxième année de premier cycle, parce qu'aucun besoin potentiel en matière de simulation qui porterait sur la première année n'a été identifié, et parce que l'on peut considérer que « les cartes sont complètement redistribuées » par le *numerus clausus* et sa répartition régionale. La modélisation de la première année aurait toutefois pu être nécessaire si le *numerus clausus* avait été corrélé au nombre d'étudiants inscrits en première année, ce qui n'est actuellement pas le cas.

Pour la modélisation des études médicales, d'autres choix auraient pu être faits : par exemple, tout en conservant la modélisation des ECN, il était possible de passer directement du *numerus clausus* au nombre d'étudiants inscrits en DCEM 4, en appliquant un coefficient estimé à partir de données observées. Le déroulement des études médicales est simulé année d'études par année d'études, tant pour le PCEM/DCEM que pour le TCEM, pour pouvoir faire apparaître les effets d'un changement des comportements des étudiants en médecine (en matière de redoublements ou d'abandons des études médicales par exemple).

Dans la version précédente du modèle, le déroulement des études médicales était simulé année d'études par année d'études également pour les deux premiers cycles. Pour le troisième cycle en revanche, le nombre de diplômés était déduit directement du nombre d'entrants en troisième cycle.

Une fois le **nombre de diplômés** calculé, il est nécessaire de déterminer le nombre de médecins qui exerceront la médecine parmi eux, c'est-à-dire de calculer le nombre de « fuites » (voir C. 1. a.), sauf si l'on fait l'hypothèse que tous les jeunes diplômés exercent la médecine en France, hypothèse qui n'a pas été retenue, la proportion des jeunes diplômés n'exerçant jamais la médecine en France ayant 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).

La cessation d'activité et la reprise d'activité n'étaient pas modélisées dans la version précédente du modèle, le champ incluant les médecins en cessation d'activité.

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. Au cours des études

préliminaires, leur fréquence a été évaluée. Ces changements ne sont pas particulièrement courants. Par exemple, au cours des trois dernières années, en moyenne, 1,4 % des médecins en activité au cours de deux années consécutives ont changé de région entre les deux années. Toutefois, le solde migratoire a tendance à être plutôt toujours négatif ou plutôt toujours positif pour une région donnée. Les effets cumulés des changements sont donc parfois finalement importants sur le long terme. Les attentes à l'égard des résultats produits par le modèle en ce qui concerne la précision des effectifs projetés par région, par mode d'exercice et par zone d'exercice ont justifié que l'on modélise ces changements.

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, permettant ainsi d'étudier séparément les flux correspondant aux départs à la retraite et aux décès dans l'analyse des résultats des simulations.

Les cessations définitives d'activité ne correspondant pas à un départ à la retraite 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 en matière de simulation : on 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.

Les flux internationaux de médecins étant supposés nuls (voir A. 1. b. et A. 3.), les entrées des médecins diplômés à l'étranger et les sorties des médecins partant exercer à l'étranger en cours de vie active n'ont pas été modélisées.

#### ***d. Les méthodes : modélisation agrégée et microsimulation***

Pour la version précédente du modèle, la microsimulation avait été employée pour simuler l'entrée dans la vie active des jeunes diplômés et le déroulement de la carrière des médecins, tandis que la modélisation agrégée (ou modélisation par composantes) avait été employée pour modéliser le déroulement des études médicales.

La **microsimulation** est une méthode permettant de simuler l'évolution d'une population à un niveau individuel. Les calculs sont donc réalisés au niveau de chaque médecin : le point de départ de la simulation est une base de données comprenant des informations individuelles sur les médecins en activité au début de la période de projection. Ces médecins vieillissent, changent de région d'exercice, partent à la retraite, décèdent, etc., tandis que de nouveaux médecins entrent dans la vie active et viennent ainsi s'ajouter à la base initiale. Le résultat de la simulation est également une base de données individuelles.

Les résultats attendus du modèle de projections sont des effectifs de médecins actifs au cours de chaque année de la période de projection, ventilés selon certains critères (sexe, âge, spécialité, région d'exercice, mode d'exercice, et zone d'exercice), mais il doit être possible de ventiler ces effectifs selon plusieurs critères croisés (même si, du fait de la petite taille des effectifs ainsi obtenus, tous les croisements ne sont pas sensés). Compte tenu du niveau de détail des effectifs à projeter, mais aussi du nombre et du type « d'événements » à modéliser, l'utilisation de la modélisation agrégée pour la construction du modèle supposait l'utilisation de formules complexes et/ou de matrices (de paramètres) de grandes tailles. Il aurait peut-être été nécessaire, compte tenu de la complexité des calculs, de réduire le niveau de détail des résultats si la modélisation agrégée avait été utilisée pour modéliser toute la carrière des médecins. La complexité des calculs aurait encore été accrue si l'on avait souhaité tenir compte de l'hétérogénéité des comportements de médecins, c'est-à-dire en

pratique si les probabilités de vivre un événement donné (le départ à la retraite par exemple) étaient déclinées selon les caractéristiques des médecins : plus le nombre de caractéristiques retenues aurait été élevé, plus les calculs auraient été complexes. Il aurait alors sans doute fallu faire des hypothèses restrictives quant à l'hétérogénéité de la population des médecins : leurs comportements auraient été supposés plus homogènes, afin de réduire le nombre de paramètres.

Finalement, la microsimulation a été retenue comme méthode pour les raisons suivantes :

- par souci de confort dans la programmation, mais aussi dans l'utilisation et la présentation du modèle
- afin de permettre au besoin de prendre en compte l'hétérogénéité de la population des médecins en ce qui concerne leurs comportements
- pour garantir un niveau de détail des résultats suffisant
- pour la souplesse que procure cette méthode : ajouter de nouveaux paramètres ou modéliser de nouveaux événements est alors relativement simple.

Notons que les trajectoires individuelles projetées des médecins ne feront pas l'objet d'analyses. Seuls les effectifs projetés (selon les différentes caractéristiques) au cours de chaque année seront étudiés. Par exemple, peu importe que le médecin partant à la retraite entre 2025 et 2026 soit celui ayant repris une activité en 2020 ou celui n'ayant jamais interrompu son activité. Nous nous intéresserons à l'effectif de médecins actifs.

Pour la simulation du déroulement des études médicales, ont été utilisées la modélisation agrégée pour le premier et le deuxième cycles, et la microsimulation pour le troisième cycle. En effet, la modélisation des deux premiers cycles est assez simple et suppose seulement l'utilisation de quelques formules. Il n'est pas nécessaire que le modèle produise des informations individuelles sur les étudiants en PCEM ou en DCEM, car les informations que l'on souhaite retirer des simulations à ce stade de la carrière médicale sont réduites. De même, il n'est pas nécessaire que le modèle produise des informations individuelles sur les affectations prononcées à l'issue des ECN, ni sur les étudiants en troisième cycle et les diplômés en médecine. Par conséquent, la modélisation agrégée a été employée pour modéliser les ECN. En revanche, pour plus de confort, la microsimulation a été choisie pour modéliser le déroulement du troisième cycle. Utiliser la modélisation agrégée pour le troisième cycle aurait supposé l'utilisation de formules plus nombreuses et plus complexes.

En microsimulation, la simulation repose en grande partie sur la sélection d'individus choisis pour vivre chacun des événements modélisés. **La sélection est effectuée par un tirage aléatoire, réalisé à l'aide d'une loi uniforme** : une réalisation d'une loi uniforme (appelée « alea ») est simulée sur le champ sur lequel doit être réalisée la sélection. Les propriétés de la loi uniforme sont telles que sur l'ensemble du champ sur lequel doit être effectuée la sélection, 5 % des individus doivent se voir 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 « alea » est d'autant plus proche de cette répartition théorique que l'effectif correspondant au champ sur lequel la sélection doit être effectuée est grand. Il reste ensuite, pour chaque individu, à comparer la valeur de la variable « alea » obtenue à la probabilité pour cet individu de vivre l'événement en question, estimée préalablement (en fonction éventuellement des caractéristiques de l'individu). Les individus pour lesquels la valeur de « alea » est inférieure à la probabilité pour l'individu de vivre l'événement sont sélectionnés pour vivre l'événement (voir tableau 1).

Tableau 1 - Exemple de la simulation du départ à la retraite

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	2	Homme	Généraliste
7	Femme	Généraliste	3	Homme	Généraliste
8	Femme	Généraliste	4	Homme	Spécialiste
9	Femme	Généraliste	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.

Pour simuler la réalisation d'une loi uniforme à l'aide de la procédure ranuni de SAS (le logiciel utilisé pour la construction du modèle), il faut **choisir une « racine »**, c'est-à-dire un nombre quelconque, positif (non strictement), qui initialise l'algorithme. Deux simulations successives de la réalisation d'une loi uniforme réalisées avec la racine zéro ne donnent pas le même résultat, l'ordinateur utilisant dans ce cas la date machine pour initialiser l'algorithme. Or, il est préférable de pouvoir retrouver les mêmes résultats si les mêmes programmes sont lancés plusieurs fois (pour refaire par exemple la simulation d'un scénario déjà réalisé).

Par conséquent, il n'est possible de retenir comme racine qu'un nombre strictement positif. La simulation de la réalisation d'une loi uniforme donne alors toujours le même résultat (la succession des valeurs générées dans l'ordre du fichier est toujours la même), ce qui signifie que si les enregistrements du fichier dans lequel cette simulation est réalisée sont triés dans le même ordre, il est probable que les mêmes individus seront sélectionnés pour vivre différents événements modélisés. Cela peut se produire par exemple si les probabilités sont déclinées pour deux événements successifs selon les mêmes caractéristiques (ou critères). Dans ce cas, le fichier sera trié selon ces critères au moment de la simulation du premier événement, pour y intégrer les probabilités de vivre cet événement. Il sera trié dans le même ordre au moment de la simulation du deuxième événement, dans le même but, et les individus sélectionnés pour vivre le premier événement modélisé auront ainsi plus de « chances » d'être sélectionnés pour vivre le deuxième événement. Afin d'éviter d'introduire ainsi

des corrélations fictives, la racine est modifiée pour chaque événement modélisé. Mais il faut alors changer également de racine pour chaque année, sinon pour un événement donné, la même réalisation de la loi uniforme sera simulée chaque année.

Le résultat de la simulation d'un scénario est dépendant des différentes réalisations de la loi uniforme simulées. Ce résultat est donc aléatoire. Il est d'usage, pour réduire l'effet de l'aléa sur les résultats produits, de réaliser plusieurs simulations d'un scénario (itérations) avec des racines différentes, et de **présenter les résultats moyens des différentes itérations**. Afin de déterminer le nombre d'itérations à réaliser, les résultats de tirages aléatoires correspondant à des probabilités plus ou moins fortes (0,01 ; 0,05 ; 0,1 ; 0,5) effectués avec différentes racines ont été étudiés, de même que la stabilité des résultats moyens en fonction du nombre d'itérations réalisées. La stabilité des effectifs projetés de médecins en fonction du nombre d'itérations réalisées a également été étudiée. Chaque itération allongeant le temps de calcul, l'objectif était de déterminer le nombre d'itérations le plus faible possible qui permettait de réduire l'effet de l'aléa de façon satisfaisante. Au vu des résultats des différents tests effectués, nous avons fixé le nombre d'itérations à réaliser pour chaque scénario à 5.

Dans la version précédente du modèle, 10 itérations étaient réalisées pour chaque scénario, et la racine n'était pas modifiée pour chaque événement et chaque année de la période de projection.

Finalement, la racine choisie pour la simulation de la réalisation de la loi uniforme pour un événement donné, une année donnée de la période de projection, et pour une itération donnée d'un scénario est :

« iteration »+ « annee »+ « eve » où

« iteration » est la valeur retenue pour initialiser l'algorithme pour l'itération en question,

« annee » l'année courante de la période de projection

« eve » le numéro de l'événement dans la liste chronologique des événements modélisés.

## ***2. Le choix des sources***

### ***a. Les sources de données sur les médecins actifs : ADELI et les fichiers de l'Ordre***

Deux bases de données pouvaient être utilisées comme base initiale pour les projections : le répertoire ADELI et le fichier annuel de l'Ordre des médecins.

Pour la **version précédente du modèle**, c'est le répertoire **ADELI** qui était utilisé. Le choix de la base initiale conditionne en partie la nature des résultats des projections. Ainsi, il n'est pas possible de distinguer les médecins en cessation temporaire d'activité à partir du répertoire ADELI. S'il est utilisé comme base initiale pour les projections, les effectifs projetés les incluent donc. Retenir comme base initiale le fichier de l'Ordre implique de projeter les effectifs par spécialité en retenant la définition de chaque spécialité de l'Ordre. Mais le choix de la base initiale a aussi des conséquences sur la construction du modèle. Ainsi, la notion de mode d'exercice dans les fichiers de l'Ordre est proche de celle d'ADELI, mais elle est différente. Il n'est donc pas possible, si l'on utilise ADELI comme base initiale, de modéliser les changements de mode d'exercice, puisque pour ce faire, il faut calculer les probabilités de changer de mode d'exercice à partir des fichiers de l'Ordre (voir ci-après), et que le mode d'exercice des fichiers de l'Ordre ne correspond pas à celui d'ADELI.

Pour la **version 2008** du modèle de projections, nous avons retenu le **fichier annuel de l'Ordre** (au 31/12/2006) comme base initiale, pour plusieurs raisons, dont voici les principales :

- ADELI doit disparaître, et le **RPPS** (répertoire partagé des professions de santé) être mis en place prochainement. Or, le RPPS sera principalement alimenté par les données de l'Ordre.

Son contenu devrait par conséquent être plus proche des fichiers de l'Ordre actuels que d'ADELI, en ce qui concerne les concepts. Utiliser le fichier annuel de l'Ordre permet donc d'assurer une certaine pérennité à la version révisée du modèle, et d'éviter des ruptures entre les projections de 2008 et les suivantes.

- L'enregistrement des nouveaux entrants d'une part et des départs à la retraite et des décès d'autre part est effectué plus systématiquement dans les fichiers de l'Ordre que dans le répertoire ADELI : les effectifs correspondant aux médecins les plus âgés sont plus élevés dans ADELI que dans les fichiers de l'Ordre, contrairement aux effectifs correspondant aux médecins les plus jeunes. De plus, les effectifs de médecins salariés tirés d'ADELI sont sous-estimés, et par conséquent corrigés (un redressement est effectué à partir des données de l'Ordre). Toutefois, pour certaines spécialités, les effectifs restent sous-estimés après corrections (c'est le cas de l'anesthésie-réanimation par exemple<sup>1</sup>).
- L'identifiant des médecins dans ADELI n'est pas toujours pérenne : lorsque les médecins changent de situation, ils changent parfois également d'identifiant. De ce fait, il est difficile d'étudier les changements de région, de mode ou de zone d'exercice à partir d'ADELI (par appariement de fichiers annuels successifs), et de calculer des probabilités de changer de situation à partir de cette source. Pour ce faire, il faut donc calculer les probabilités à partir des données de l'Ordre. Retenir le fichier annuel de l'Ordre comme base de données individuelles permet d'assurer une certaine cohérence entre les données relatives au stock des médecins en activité au début de la période de projection et les flux estimés.
- Utiliser comme base initiale le fichier annuel de l'Ordre permet de projeter des effectifs de médecins en activité régulière (en excluant les retraités ayant conservé ou repris une petite activité) ou remplaçants, en excluant les médecins ayant cessé temporairement leur activité.

Le fichier annuel de l'Ordre présente toutefois des caractéristiques gênantes pour la construction du modèle de projections :

- Les **remplaçants n'ont pas de mode d'exercice** dans ce fichier. On ne sait pas non plus s'ils remplacent des médecins libéraux, salariés hospitaliers, etc. Quatre modalités ont été retenues pour la variable mode d'exercice (voir A. 1. a.). Deux solutions étaient envisageables : créer une cinquième modalité pour la variable mode d'exercice (« remplaçant ») ou **imputer** aux remplaçants l'une des quatre modalités existantes. C'est cette deuxième solution qui a été retenue. Le nombre de remplaçants était trop faible pour constituer une cinquième modalité, pour laquelle les effectifs projetés auraient pu ne pas être suffisamment fiables. Par ailleurs, il aurait fallu, si les remplaçants étaient distingués des médecins en activité régulière, simuler le passage du statut de remplaçant à celui de médecin en activité régulière, et vice-versa. Enfin, du point de vue de l'analyse, il aurait été intéressant de disposer des effectifs projetés de remplaçants, mais ces effectifs sont nettement moins intéressants dès lors qu'ils ne peuvent pas être déclinés par mode d'exercice.

Une troisième solution aurait consisté à ôter les remplaçants du champ des projections, solution qui n'a pas été retenue pour éviter de sous-estimer l'effectif de médecins actifs. Cette solution aurait également impliqué la modélisation du passage du statut de remplaçant à celui de médecin en activité régulière et vice-versa.

Pour imputer aux remplaçants un mode d'exercice, des probabilités ont été estimées à partir des données ADELI au 01/01/2007, qui donnent le mode d'exercice de l'activité principale des médecins en activité, y compris les remplaçants, que l'on peut repérer. Ces probabilités ont été déclinées selon le sexe des médecins et le fait qu'ils étaient généralistes ou spécialistes. Les effectifs de médecins remplaçants salariés répertoriés dans ADELI au 01/07/2007 étaient extrêmement faibles ; ils ont donc été regroupés avec les médecins remplaçants dont l'activité

---

<sup>1</sup> Voir Rapport de l'Observatoire national de la démographie des professions de santé 2005, Tome 3.

était mixte. Finalement, les modalités « libéral » ou « mixte » ont été attribuées aux remplaçants du fichier de l'Ordre pour la variable mode d'exercice.

L'estimation de ces probabilités à partir d'ADELI a été préférée à une autre méthode d'estimation, plus approximative, qui aurait consisté à observer le mode d'exercice au cours de l'année  $n+1$  des remplaçants qui deviennent actifs réguliers entre l'année  $n$  et l'année  $n+1$ , et à faire l'hypothèse que les remplaçants ne changent pas de mode d'exercice en quittant ce statut.

L'estimation des probabilités à partir des données d'ADELI a semblé moins approximative, mais les probabilités ainsi estimées ne pourront être actualisées, ADELI devant disparaître.

- La commune d'exercice du médecin, de laquelle nous déduisons la zone d'exercice, n'est pas indiquée pour environ 1,8 % des médecins en activité régulière exerçant en France métropolitaine (hors Corse). Le fichier annuel de l'Ordre ne donne pas non plus cette information pour les remplaçants. Or, les effectifs de médecins doivent être projetés par zone d'exercice pour la France métropolitaine. **L'information sur la zone d'exercice n'est donc pas nécessaire pour les médecins exerçant dans les DOM-TOM et en Corse, mais elle doit être imputée pour les médecins en activité régulière exerçant en métropole (hors Corse) pour lesquels elle n'est pas indiquée, ainsi que pour les remplaçants.**

Pour imputer la zone d'exercice, des probabilités déclinées par région et selon que les médecins sont généralistes ou spécialistes ont été calculées à partir des informations sur les médecins pour lesquels le code commune (donc la zone d'exercice) était connu. A défaut d'information spécifique aux remplaçants, on suppose ainsi que le comportement des remplaçants en ce qui concerne le choix de la zone d'exercice est le même que celui des médecins en activité régulière. Il en est de même pour les médecins actifs pour lesquels le code de la commune d'exercice n'est pas indiqué. Par conséquent, la répartition des médecins par zone d'exercice au sein de chaque strate composée par la région et le fait d'être généraliste ou spécialiste n'est pas déformée par l'imputation.

Avant d'imputer le mode d'exercice ou la zone d'exercice, une recherche est effectuée dans le fichier de l'année précédente. Si pour le même médecin, l'information sur le mode d'exercice ou sur la zone d'exercice était indiquée dans le fichier précédent, elle est dupliquée pour l'année suivante (même si dans les faits, le médecin peut avoir changé de mode d'exercice ou de zone d'exercice).

Le mode d'exercice et la zone d'exercice ne sont pas imputés pour les médecins en cessation temporaire d'activité. Ils sont affectés au cours de la simulation à ceux de ces médecins auxquels le modèle fait reprendre une activité au cours de la période de projection.

Quelques autres redressements, concernant peu de médecins, ont été effectués sur le fichier de l'Ordre au 31/12/2006 :

- Les médecins âgés de plus de 80 ans (une centaine) ont été ôtés de la base initiale, afin que le champ de la base initiale soit le même que celui des effectifs projetés (voir A. 1. b.).
- Pour environ 250 médecins actifs, le mode d'exercice était inconnu. Les médecins concernés étant peu nombreux, le mode libéral a été attribué aux généralistes et le mode salarié hospitalier aux spécialistes.
- Pour environ 1 100 médecins (soit environ 0,5 %) des médecins actifs, l'année d'entrée dans la vie active était manquante ou incorrecte (antérieure à l'année d'obtention du diplôme). La date d'obtention du diplôme a été attribuée comme date d'entrée dans la vie active aux médecins concernés.

Ces différents redressements ont été également effectués pour les autres fichiers annuels de l'Ordre utilisés par le modèle (pour l'estimation des probabilités), et non seulement pour le dernier en date, qui a servi à constituer la base de données individuelles initiale.

Le fichier annuel de l'Ordre au 31/12/2006 a permis de constituer la base initiale des médecins en activité. Les fichiers annuels de l'Ordre au 31/12/2003, au 31/12/2004, au 31/12/2005 et au 31/12/2006 ont également été retenus comme source pour **l'estimation des probabilités** servant aux modules 2 (entrée dans la vie active), 3 (changements en cours de vie active) et 4 (sortie de la vie active). Il s'agissait d'assurer une certaine cohérence entre les données utilisées pour constituer la base initiale et les données utilisées pour l'estimation des probabilités.

Les tables de mortalité par PCS de l'Insee ont été utilisées pour modéliser les décès (voir E. 2.).

### ***b. Les sources de données sur les étudiants en médecine***

Deux sources principales de données mobilisables ont été identifiées :

- les fichiers SISE (système d'information sur le suivi de l'étudiant) de la DEPP, contenant des informations individuelles et annuelles sur les étudiants inscrits à l'université (depuis l'année universitaire 1994-1995), et les résultats obtenus aux diplômes (depuis l'année universitaire 1999-2000 pour les résultats aux diplômes de santé) ;
- les fichiers des résultats individuels aux ECN, mises en place en 2004.

Il existe cependant d'autres sources de données sur les étudiants en médecine et les diplômes de médecine délivrés :

- L'ONDPS réalise chaque année depuis 1999 (année universitaire 1998-1999) un recensement des diplômes d'études spécialisées (DES) de chaque spécialité délivrés par les facultés de chaque région, mais ce recensement ne concerne pas les diplômes de médecine générale. Les résultats de ces recensements sont publiés annuellement sur le site de l'ONDPS.
- La DEPP a réalisé jusqu'à l'année universitaire 2000-2001 une enquête, appelée enquête n° 13, auprès de toutes les facultés de médecine, invitées à indiquer les nombres de diplômes de docteur en médecine délivrés par spécialité (médecine générale incluse), en distinguant les étudiants selon leur sexe et selon leur nationalité (nationalité française / nationalité étrangère).
- Des données ont été produites par l'AUFEMO, portant sur les effectifs inscrits et le nombre de redoublants par faculté en quatrième année de DCEM pour les années 2004-2005 et 2005-2006.

Les données issues de chacune de ces sources ont été étudiées afin d'évaluer leur fiabilité, et d'en déterminer les utilisations possibles.

### **SISE**

Les fichiers SISE des inscriptions pour les années 2000-2001 à 2006-2007 (dernière année disponible pour les inscriptions) et des diplômes pour les années 2000-2001 à 2004-2005 (dernière année disponible pour les diplômes) ont été exploités.

Les données issues de SISE ont permis de calculer les effectifs d'étudiants inscrits à la préparation d'un diplôme de docteur en médecine par niveau d'études, par faculté (donc par région) et par diplôme (spécialité) des années 2000-2001 à 2006-2007. Ont également été calculés à partir de ces données des taux de dépassement du *numerus clausus* (proportions d'étudiants poursuivant des études médicales à l'issue de la première année en plus du *numerus clausus* fixé), des taux de redoublements, des taux d'abandons\* (incluant les décès, voir ci-dessous), des taux d'arrivées\*\* (voir ci-dessous) et des taux de mobilité interrégionale, par niveau d'études et par région (et par spécialité pour le troisième cycle), grâce à l'appariement de fichiers annuels deux à deux.

\* Le taux d'abandons ne rend pas compte uniquement de l'abandon des études médicales à proprement parler. Il s'agit de la proportion d'étudiants présents une année  $n$  dans le fichier des



inscrits dans le cursus médical que l'on ne retrouve pas inscrits dans le cursus médical l'année  $n+1$ . Le taux d'abandons inclut donc les réorientations (vers une autre discipline que la médecine), la mortalité, les interruptions d'études (pour maladie, grossesse, convenances personnelles par exemple), les départs à l'étranger en cours d'études, mais il est biaisé par les réimmatriculations (un étudiant changeant d'identifiant entre l'année  $n$  et l'année  $n+1$  n'est pas retrouvé). Le biais du taux d'abandons lié aux réimmatriculations est compensé par

\*\* le taux d'arrivées puisqu'un étudiant réimmatriculé disparaît sous un identifiant (il est alors comptabilisé dans les abandons) mais réapparaît sous un autre (il est cette fois comptabilisé parmi les arrivées) lorsqu'on apparie les fichiers de deux années successives. Les étudiants qui ont interrompu leurs études pendant une ou plusieurs années universitaires ont eux aussi été comptabilisés l'année de leur interruption parmi les abandons, mais font partie des arrivées l'année de leur reprise d'études. Les arrivées incluent également les arrivées d'étudiants étrangers en cours de cursus.

Le nombre de diplômes délivrés par région et par spécialité pour chacune des années de 2000-2001 à 2004-2005 a également été calculé.

***\* Concernant les effectifs d'inscrits en PCEM et en DCEM et les taux de dépassement :***

Les divers contrôles effectués sur les données SISE ont fait apparaître que ces données sont globalement de bonne qualité.

Les taux de dépassement calculés laissent penser que les effectifs en PCEM 2 sont dans l'ensemble plutôt corrects. La comparaison des effectifs inscrits en DCEM 4 et des effectifs de candidats aux ECN issus du DCEM 4 montre pour les années 2004-2005 et 2005-2006 une concordance parfaite ou quasi-parfaite pour la grande majorité des régions.

Pour la région Nord-Pas-de-Calais, les effectifs d'inscrits à la faculté catholique de Lille doivent être ajoutés à ceux de l'université. Pour les régions Rhône-Alpes et Aquitaine, les étudiants des écoles du Service de santé des armées de Bordeaux et de Lyon sont vraisemblablement inclus dans les effectifs d'inscrits en faculté. Or, les médecins militaires ne sont pas inscrits à l'Ordre. Il est donc préférable de ne pas inclure dans la simulation les futurs médecins militaires aux effectifs futurs de médecins en activité, afin de ne pas créer de rupture de série. Puisqu'il n'est pas possible de repérer dans les données SISE les étudiants du Service de santé des Armées, la correction a porté sur le taux de dépassement. Les taux de dépassement utilisés en entrée ont été ramenés à 2,6 % (moyenne des taux de dépassement des autres régions).

Toutefois, quelques anomalies ont été repérées et ont fait l'objet de corrections.

Les effectifs d'inscrits en 2006-2007 correspondant à chaque niveau (du PCEM 2 au DCEM 4) et à chaque région seront utilisés en entrée du modèle.

***\* Concernant les effectifs d'inscrits en TCEM :***

Les contrôles effectués sur les données relatives aux inscriptions en troisième cycle ont fait apparaître de légères incohérences dans les données. Des étudiants sont inscrits à des niveaux d'études qui n'existent pas pour leur discipline (niveau 05 pour la biologie médicale, la médecine du travail, la psychiatrie et la gynécologie médicale), mais les effectifs concernés sont très faibles. Les enregistrements en question ont été redressés. Les effectifs d'inscrits en médecine du travail semblent également assez fluctuants. Mais l'analyse des différentes répartitions a surtout montré la faiblesse des effectifs calculés par discipline et par niveau d'études au niveau régional, qui a incité à la prudence pour la construction de la partie du modèle correspondant au TCEM.

Les effectifs d'inscrits en 2006-2007 correspondant à chaque niveau du TCEM, à chaque discipline et à chaque région sont utilisés en entrée du modèle.

***\* Concernant les taux de redoublements, d'abandons, d'arrivées et de changements de région pour le PCEM et le DCEM :***

La comparaison, sur deux années, pour le niveau DCEM 4, des taux de redoublements obtenus par l'appariement des fichiers SISE des inscriptions et des taux obtenus à partir des données de l'AUFEMO laisse penser que les taux obtenus par l'appariement des fichiers SISE des inscriptions sont dans l'ensemble de bonne qualité.

Quelques taux de redoublements et d'abandons anormalement élevés ont été repérés et redressés. Ils sont le plus souvent liés aux anomalies détectées sur les effectifs d'inscrits. L'évolution des taux d'abandons et de redoublements d'une année à la suivante est en général assez peu heurtée, mais la qualité des fichiers SISE (notamment la qualité de l'identifiant permettant les appariements, donc le repérage des redoublements et abandons) est décroissante lorsqu'on remonte dans le passé.

Les proportions d'étudiants ayant changé de région en cours de PCEM ou de DCEM sont très faibles. Au niveau régional, pour chaque année, tous niveaux confondus, elles sont de l'ordre de 0 à 2 %. Toutefois, les changements de région peuvent occasionner assez facilement des réimmatriculations, qui empêchent justement le repérage des changements de région. Les taux obtenus sont donc probablement un peu sous-estimés.

Pour un certain nombre d'étudiants, du fait de mauvais codages (du niveau d'études en particulier), l'évolution d'une année sur l'autre est « anormale ». Par exemple, on peut trouver quelques étudiants passant du PCEM 2 au DCEM 2 d'une année à la suivante. Les effectifs concernés sont cependant très faibles.

Les taux de redoublements, d'abandons et d'arrivées moyens (calculés sur un nombre d'années paramétrable) sont utilisés en entrée du modèle.

***\* Concernant les taux de redoublements, d'abandons, d'arrivées, de poursuites d'études (en DESC) et de changements de région pour le TCEM***

Les différents taux observés sont plutôt plus élevés que les mêmes taux calculés pour le PCEM et le DCEM. Toutefois, leurs valeurs restent dans l'ensemble relativement faibles. Certaines d'entre elles ont cependant du être redressées. L'évolution des taux d'abandons et de redoublements d'une année à la suivante est parfois un peu heurtée, sans que l'observation des effectifs correspondants conduise toutefois à douter des taux obtenus.

Les taux de redoublements, d'abandons et d'arrivées moyens (calculés sur un nombre d'années paramétrable) sont utilisés en entrée du modèle.

Finalement, les données SISE relatives aux inscriptions semblent fiables, bien que de qualité variable selon les universités, et utilisables (moyennant quelques corrections préalables et quelques précautions d'emploi) pour la modélisation du cursus des études médicales.

***\* Concernant les effectifs de diplômés***

En revanche, la qualité des données issues de SISE sur les diplômes délivrés pour les années 2000-2001 à 2004-2005 est moindre.

Si l'on en croit la comparaison du nombre de diplômés dans une discipline ECN donnée une année donnée et le nombre de postes ouverts à l'internat 3, 4 ou 5 ans auparavant (selon les disciplines), le nombre de diplômes délivrés est nettement sous-estimé, même pour les années les plus récentes. Si l'on fait abstraction des redoublements, abandons et arrivées, qui ne sont pas particulièrement fréquents et qui se compensent partiellement, les effectifs de diplômés doivent être en effet proches des nombres de postes ouverts quelques années plus tôt. Or, par exemple, pour la gynécologie-obstétrique, on compte 110 diplômés en 2004-2005 alors que cinq ans plus tôt, en 2000, 140 postes étaient ouverts à l'internat dans cette discipline.

Des discussions avec l'AUFEMO et avec la DEPP ont confirmé ces observations.

Ces effectifs de diplômés ne sont donc pas utilisés pour calculer des taux de diplômés, que l'on pourrait obtenir en rapportant les effectifs de diplômés d'une année  $n$  aux effectifs d'inscrits en année terminale d'un DES la même année.

Pour plusieurs années et pour chaque spécialité (hors médecine générale) le nombre de diplômés obtenus à partir des fichiers SISE des diplômés a été comparé au nombre de diplômes délivrés recensés chaque année par l'ONDPS. Cette comparaison montre un défaut de concordance assez net entre les deux sources. Toutefois, les répartitions (au niveau national) par spécialité des diplômés des disciplines ECN « spécialités médicales » d'une part et « spécialités chirurgicales » d'autre part calculées à partir des données de SISE sur les diplômés sont proches de celles calculées à partir des données de l'ONDPS. Les répartitions calculées à partir des données de SISE ont donc été utilisées en entrée du modèle.

### **Les données du CNG sur les ECN**

Les données sur les ECN sont issues des fichiers de gestion des ECN, qui servent notamment à l'affectation des étudiants à l'issue des ECN. L'exploitation faite de ces fichiers à l'occasion des études sur les affectations à l'issue des quatre premières éditions des épreuves (2004 à 2007) n'a fait apparaître que des anomalies rarissimes et tout à fait négligeables.

### **L'enquête n°13 de la DEPP**

L'enquête n°13 de la DEPP donne le nombre de diplômes d'Etat de docteur en médecine délivrés par chaque faculté, et parmi ceux-ci, le nombre de femmes de nationalité française, le nombre d'hommes de nationalité française, le nombre de femmes de nationalité étrangère et le nombre d'hommes de nationalité étrangère. Cette enquête était utilisée pour extraire des données utilisées en entrée du modèle de projections d'effectifs de médecins, dans sa version précédente. La qualité de l'enquête n°13 était alors jugée imparfaite, mais satisfaisante. Cependant, la dernière enquête n°13 a été réalisée en 2000-2001 et ne sera plus reconduite. A partir de cette enquête, il n'est donc possible d'estimer un taux de diplômés<sup>2</sup> que pour 2000-2001 et les années antérieures. Or, plus on remonte dans le temps, moins les données SISE sur les inscriptions (qui servent à calculer le dénominateur du taux de diplômés) sont fiables. En outre, les données issues de l'enquête n°13 sont déjà anciennes et ne pourront être actualisées par la suite. Un problème se pose enfin pour les disciplines « spécialités médicales » et « spécialités chirurgicales ». En effet, pour ces deux disciplines, le diplôme est obtenu à la fin de la quatrième ou de la cinquième année selon la spécialité. Le modèle doit donc pouvoir diplômer un certain nombre d'inscrits en quatrième année et un certain nombre d'inscrits en cinquième année de chaque année de la période de projection. Pour cela, il faut pouvoir appliquer un taux de diplômés aux inscrits en quatrième année et un taux de diplômés aux inscrits en cinquième année, donc connaître, pour les années passées, le nombre de diplômés par spécialité pour distinguer les

---

<sup>2</sup> L'estimation d'un taux de diplômés est nécessaire si l'on souhaite déterminer les effectifs de diplômés en appliquant ce taux de diplômés estimé aux effectifs d'inscrits sur les niveaux terminaux.

diplômés à l'issue de la quatrième année des diplômés à l'issue de la cinquième année. Or, l'enquête n°13 recense un nombre important de diplômés dans ces deux disciplines pour lesquels la spécialité obtenue n'est pas indiquée. Les taux de diplômés ne peuvent donc pas tous être calculés correctement. Il n'a donc pas semblé judicieux d'utiliser cette enquête comme source de données pour calculer des taux de diplômés.

Les effectifs de diplômés que l'on obtient à partir de l'enquête n°13 ont pu, pour deux années (l'année 1999 et l'année 2000), être comparés à ceux que obtenus avec les données SISE sur les diplômés. Cette comparaison établit que les effectifs obtenus à partir des fichiers SISE sont sous-estimés.

### **Le recensement annuel des DES délivrés réalisé par l'ONDPS**

Ces données sont collectées annuellement depuis 1999 par l'ONDPS auprès des facultés de médecine. Les dernières données disponibles sont celles de l'année 2005, mais au moment où nous avons étudié ces données, la dernière année disponible était 2004.

Ce recensement ne concerne pas la médecine générale, ce qui est un inconvénient.

Les comparaisons avec les données issues de SISE et de l'enquête n°13 ont fait apparaître des écarts. Les écarts avec SISE ne sont pas révélateurs d'un défaut des données de l'ONDPS, puisque les données de SISE sur les diplômés ne sont pas parfaites. En revanche, les écarts constatés avec l'enquête n°13 pour les années 1999 et 2000 posent davantage question, mais nous n'avons aucune raison de penser *a priori* que l'une de ces deux sources est plus fiable que l'autre.

On constate une rupture de série assez nette dans les données de l'ONDPS en 2003 : le nombre total de diplômés (hors médecine générale) est de 2 005 en 2002, de 1 595 en 2003 et de 2 292 en 2004. L'effectif total obtenu pour 2003 est peu plausible. La comparaison avec les données issues de l'enquête n°13 fait apparaître un nombre important de diplômés sans spécialité précisée pour les disciplines « spécialités médicales » et « spécialités chirurgicales » qui semble ne pas avoir de correspondance dans les effectifs de diplômés recensés par l'ONDPS.

Les données collectées par l'ONDPS n'ont donc pas non plus été utilisées estimer des taux de diplômés : compte tenu des effectifs faibles pour certaines disciplines, les taux de diplômés estimés sont assez sensibles, voire très sensibles aux erreurs éventuelles sur les effectifs de diplômés.

### ***3. Le choix des hypothèses***

Les hypothèses retenues pour les simulations diffèrent d'un scénario à l'autre, puisque les variantes visent justement à évaluer l'impact du changement d'une hypothèse, correspondant au changement d'un comportement des médecins ou à la mise en place d'une mesure de régulation.

Toutefois, l'une des hypothèses est centrale. Le scénario tendanciel, qui sert de référence, notamment, est basé sur l'**hypothèse de comportements des médecins constants**, c'est-à-dire identiques à ceux observés actuellement. Ainsi, les probabilités pour les médecins de vivre un événement donné sont-elles calculées à partir de données observées au cours de la période récente, c'est-à-dire en observant les comportements actuels des médecins.

En particulier, les probabilités de choisir telle ou telle région d'exercice, tel ou tel mode d'exercice et telle ou telle zone d'exercice, ainsi que les probabilités de changer de région, de mode ou de zone d'exercice en cours de carrière sont supposées égales à celles observées au cours des dernières années. On fait donc l'hypothèse que les éventuelles opportunités qui pourraient se présenter (ou au contraire

ne plus se présenter) aux médecins à l'avenir, du fait des départs massifs à la retraite dans certaines régions (en particulier dans les régions « attractives »), ou de l'ouverture (ou la fermeture) de postes hospitaliers par exemple ne modifieront pas les comportements des médecins quant au choix de la région, du mode et de la zone d'exercice.

L'hypothèse de comportements constants n'est pas la seule hypothèse possible en matière de comportements des médecins. On aurait pu, par exemple, retenir l'hypothèse d'une prolongation des tendances observées récemment. Prenons l'exemple du choix de la spécialité selon le sexe. En faisant l'hypothèse de comportements constants en la matière, nous supposons que les nouveaux entrants dans la vie active au cours de la période de projection seront aussi fréquemment qu'aujourd'hui des femmes, c'est-à-dire que la tendance à la féminisation observée actuellement (au moins pour certaines spécialités) ne se poursuivrait pas à l'avenir. Il était possible de faire au contraire l'hypothèse que cette tendance à la féminisation se poursuivrait, c'est-à-dire que parmi les nouveaux entrants dans la vie active de demain, on compterait une proportion encore plus importante de femmes qu'aujourd'hui. Si l'on fait cette hypothèse, il faut toutefois être capable de dessiner la tendance future, c'est-à-dire de déterminer les proportions de femmes parmi les nouveaux entrants dans la vie active pour chacune des années à venir, éventuellement déclinées selon les critères que l'on souhaite retenir.

L'hypothèse de comportements constants n'est pas la plus probable, mais elle est la plus neutre. De plus, il n'est pas nécessaire, dans le cadre de projections, qui n'ont pas vocation à être interprétées comme des prévisions, de prévoir, c'est-à-dire de déterminer de façon aussi certaine que possible les évolutions futures des comportements des médecins.

Certaines variantes diffèrent du scénario tendanciel sur une hypothèse relative aux comportements des médecins, qui ne sont alors plus supposés constants, afin justement de mesurer l'impact d'un changement dans les comportements des médecins, spontané ou induit par une mesure de régulation.

D'autres hypothèses sont faites pour chaque simulation, par exemple sur le niveau national du *numerus clausus* et sa répartition régionale pour les différentes années de la période de projection, sur la répartition des postes ouverts aux ECN par discipline et par région, sur le rapport entre le nombre de postes ouverts aux ECN et le nombre de candidats, etc. Le plus souvent, c'est la constance par rapport à la situation actuelle qui est supposée, mais ce n'est pas toujours le cas : le niveau national du *numerus clausus*, par exemple, a été fixé pour les années à venir à un niveau différent de son niveau actuel.

Des hypothèses sont également sous-jacentes à la construction même du modèle : ainsi, dans la simulation du déroulement des études médicales, les migrations géographiques des étudiants en cours d'études (entre les différentes régions) ne sont pas prises en compte, ce qui signifie que l'on fait l'hypothèse que les étudiants en médecine ne changent pas de région au cours de leurs études, sauf éventuellement au moment de l'entrée en troisième cycle, ou plus exactement que les soldes des entrées et sorties sont nuls pour chaque région.

De même, les migrations internationales des médecins au cours de leur période d'activité ne sont pas prises en compte pour les projections. L'hypothèse sous-jacente est qu'aucun médecin diplômé en France et ayant commencé sa carrière médicale en France ne part exercer à l'étranger au cours de sa vie active d'une part, et qu'aucun médecin diplômé à l'étranger ne vient exercer en France d'autre part (ou que les flux internationaux se compensent, tant en niveau qu'en ce qui concerne leur composition).

#### ***4. L'enchaînement des calculs***

Avant que ne soit réalisée la simulation proprement dite, **les paramètres définissant le scénario** dont il s'agit de simuler la réalisation sont **pris en compte**, et les bases de données et les tables utilisées en entrée sont préparées.

Lorsque les calculs sont effectués par microsimulation, ils sont faits à partir de **deux bases de données individuelles**, qui évoluent au cours de la simulation, de sorte qu'elles représentent toujours :

- pour l'une les **médecins en activité** (Cette base comprend un enregistrement par médecin actif au cours de l'année courante de la période de projection)
- et pour l'autre les **médecins en cessation temporaire d'activité**.

À partir de données sources (voir A. 2. a.), ces deux bases de données individuelles sont donc constituées. Elles serviront de point de départ à la simulation. À ce stade, ces bases décrivent respectivement la population des médecins actifs et des médecins en cessation temporaire d'activité avant le début de la période de projection.

Sont également créées à partir de la même source des **bases de données individuelles relatives aux médecins** en activité (ou en cessation temporaire d'activité) **au cours de l'une des années précédant le début de la période de projection**. Elles **servent à estimer les différents jeux de probabilités** utilisés par le modèle pour réaliser les tirages aléatoires à l'aide desquels sera simulée la réalisation du scénario (voir A. 1. d.).

Enfin, à partir de données sources (sur les inscriptions universitaires en médecine, les ECN, la mortalité par exemple) et des bases de données individuelles décrites ci-dessus, sont préparées **les tables que le modèle utilisera en entrée** et qui contiennent pour la plupart d'entre elles des probabilités de vivre un événement donné et pour les autres des effectifs (d'étudiants au moment du début de la période de projection). Contrairement à la version précédente, la nouvelle version du modèle prépare automatiquement toutes les tables en entrée nécessaires à la simulation. Les probabilités ainsi estimées correspondent à des hypothèses spécifiques (comportements constants et identiques aux comportements observés actuellement en général), mais il est possible de les modifier pour simuler les effets d'hypothèses différentes.

Une fois les paramètres pris en compte et les bases de données et les tables utilisées en entrée préparées, le calcul des effectifs projetés commence.

Les calculs sont effectués **successivement pour chaque année de la période de projection**, mais contrairement à la version précédente du modèle, qui réalisait les calculs correspondant à une année donnée région par région, cette version du modèle effectue tous les calculs correspondant à une année donnée **pour toutes les régions simultanément**, sauf le calcul des effectifs d'étudiants en PCEM/DCEM, réalisé région par région.

Ci-dessous est décrit l'enchaînement des calculs **pour une année** donnée de la période de projection.

Dans un premier temps, sont **calculés, de manière agrégée**, les **effectifs d'étudiants en premier et deuxième cycles** des études médicales, du PCEM 1 au DCEM 4 (sous-module 1.1 – voir B. 1.), qui correspondent aux effectifs de l'année scolaire 2007-2008 pour la première année de la période de projection pour les projections de 2008, cette période commençant en 2007. Puis les **affectations prononcées à l'issue des ECN** (sous-module 1.2 – voir B. 2.) de l'année sont simulées (2007 pour la première année de la période de projection, pour les projections de 2008). Sont ensuite calculés, par microsimulation, les **effectifs d'étudiants en troisième cycle** des études médicales (sous-module 1.3 –

voir B. 3.), ainsi que le nombre de diplômés par région et par spécialité (effectifs de l'année scolaire 2007-2008 pour la première année de la période de projection, pour les projections de 2008).

À ce stade de la simulation, on dispose d'enregistrements individuels contenant des informations (la région de diplôme et la spécialité) sur les diplômés de l'année.

Sont ensuite **sélectionnés parmi ces diplômés ceux qui exerceront un jour la médecine** :  $f\%$  des enregistrements correspondant aux jeunes diplômés de l'année générés précédemment par le module 1 sont ôtés,  $f$  correspondant au taux de « fuites » et étant fixé à 3 % pour les projections de 2008 (sous-module 2.1 - voir C. 1. a.). Un délai à l'entrée dans la vie active est ensuite attribué aux jeunes diplômés qui viennent d'être sélectionnés pour exercer un jour la médecine (voir C. 1. b.).

Ensuite, **les médecins qui entrent dans la vie active au cours de l'année sont regroupés**, qu'ils aient été diplômés au cours de l'année ou au cours d'une année antérieure (voir C. 1. c. et d.).

Sont **attribués aux jeunes médecins entrant dans la vie active au cours de l'année, successivement** (voir ci-après), **un sexe et un âge simultanément** (sous-module 2.2 - voir C. 2.), **une région d'exercice** (sous-module 2.3 - voir C. 3.), **un mode d'exercice** (sous-module 2.4 - voir C. 4.) et **une zone d'exercice** (sous-module 2.5 - voir C. 5.).

À cette étape de la simulation, toutes les informations utiles sur les entrants, c'est-à-dire celles dont on dispose pour les médecins en activité, ont été générées. Cependant, les entrants ne sont pas insérés immédiatement dans la base individuelle des médecins en activité. Les enregistrements correspondants sont placés en attente, le temps de simuler dans un premier temps les cessations temporaires et définitives de la vie active, et dans un second temps les changements de région, de mode et de zone d'exercice.

Pour une année donnée de la période de projection, si les nouveaux entrants dans la vie active et les médecins reprenant une activité étaient insérés dans la base des médecins actifs avant que ne soient simulées les sorties temporaires ou définitives de la vie active, le modèle risquerait de sélectionner pour leur faire cesser leur activité des médecins qui l'ont commencée ou reprise ou cours de la même année. De plus, il semblait plus logique d'estimer des probabilités de cesser temporairement son activité et des probabilités de partir à la retraite ou de décéder sur le champ des médecins actifs au cours d'une année donnée, ce qui a été fait. Ces probabilités doivent s'appliquer au même champ : celui des médecins actifs au cours de l'année courante de la période de projection. Il s'agit alors, avant de faire entrer dans la vie active les nouveaux entrants de l'année, de sélectionner, parmi les médecins actifs au cours de l'année courante, ceux qui auront cessé leur activité à la fin de l'année, temporairement ou définitivement. Les médecins entrant dans la vie active en cours d'année sont comptabilisés, par convention, comme actifs à compter du 31 décembre de l'année, c'est-à-dire en pratique à partir de l'année suivante. Ne faisant pas partie du champ sur lequel doivent être sélectionnés les médecins cessant temporairement ou définitivement leur activité, ils doivent donc être insérés à la base des médecins actifs *a posteriori*.

Sont donc **sélectionnés simultanément** (voir ci-après) **les médecins qui cessent temporairement leur activité** (sous-module 3.4 - voir D. 4.), **ceux qui partent à la retraite** (sous-module 4.1 - voir E. 1.) et **ceux qui décèdent** (sous-module 4.2 - voir E. 2.). Les enregistrements correspondant à ces médecins sont **supprimés de la base des médecins actifs**. Ceux correspondant aux médecins qui cessent temporairement leur activité seront placés ultérieurement dans la base des médecins en cessation temporaire d'activité.

À cette étape de la simulation, la base des médecins actifs contient les enregistrements correspondant aux médecins actifs au cours de l'année courante de la période de projection et au cours de l'année suivante. Il s'agit là du champ sur lequel doivent être sélectionnés les médecins qui changent de

région, de mode ou de zone d'exercice. En effet, les probabilités de changer de région, de mode ou de zone ont été estimées sur le champ qui semblait le plus adapté : celui des médecins actifs au cours de deux années consécutives, champ auquel elles doivent donc être appliquées.

Sont donc **simulés successivement** (voir ci-après) **les changements de région** (sous-module 3.1 – voir D. 1.), **les changements de mode d'exercice** (sous-module 3.2 – voir D. 2.) et **les changements de zone d'exercice** (sous-module 3.3 – voir D. 3.).

Les changements de région sont simulés en deux étapes : dans un premier temps, les médecins changeant de région sont sélectionnés, et dans un second temps, une nouvelle région d'exercice leur est attribuée (voir D. 1.). Les changements de mode d'exercice et les changements de zone d'exercice sont simulés de la même manière.

La **mise à jour des variables région, mode et zone d'exercice**, pour les médecins en ayant changé, est effectuée seulement une fois les changements de région, de mode et de zone d'exercice simulés, afin que pour les changements de mode d'exercice (resp. de zone d'exercice), simulés après les changements de région (resp. les changements de région et les changements de mode d'exercice), le modèle travaille sur l'ancienne région d'exercice (resp. l'ancienne région d'exercice et l'ancien mode d'exercice) si celle-ci est utilisée comme critère pour l'estimation des probabilités.

Sont ensuite **sélectionnés** dans la base sur les médecins en cessation temporaire d'activité **les médecins qui reprennent une activité** (sous-module 3.5 – voir D. 5. a.). **En même temps** que sont sélectionnés les médecins reprenant une activité, sont également sélectionnés dans cette base, à l'aide de tirages simultanés (voir ci-après), **les médecins en cessation temporaire d'activité qui décèdent ou qui partent à la retraite**.

À **ceux des médecins qui reprennent une activité**, sont ensuite **attribués successivement** (voir ci-après) **un mode, une région et une zone d'exercice** (sous-module 3.5 – voir D. 5. b.).

À cette étape de la simulation, on dispose des mêmes informations sur les médecins reprenant une activité que sur médecins en activité. Les enregistrements correspondant aux médecins reprenant une activité sont placés en attente dans une table, afin d'être insérés dans la base des médecins actifs ultérieurement.

Les enregistrements correspondant aux médecins qui reprennent une activité ou à ceux qui décèdent ou qui partent à la retraite parmi ceux qui sont en cessation temporaire d'activité sont **ôtés de la base des médecins en cessation temporaire d'activité**.

Les enregistrements correspondant aux **médecins qui cessent temporairement leur activité** (sélectionnés précédemment), mis en attente, sont ensuite **placés dans la base des médecins en cessation temporaire d'activité**.

Les enregistrements correspondant aux **médecins entrant dans la vie active** et aux **médecins reprenant une activité** (sélectionnés précédemment), mis en attente, sont ensuite **insérés dans la base des médecins en activité**.

La mise à jour de la base des médecins actifs étant dès lors terminée, la **répartition des médecins actifs au cours de l'année selon toutes les variables retenues** pour les projections est calculée, et les effectifs projetés pour l'année courante sont ajoutés à la table en sortie contenant les effectifs projetés pour chaque année de la période.

Enfin, les **médecins en activité** et les **médecins en cessation temporaire d'activité** sont **vieillis d'un an**.



Il faut noter que l'ordre des différents événements est en partie (et en partie seulement) arbitraire : la sélection (et non l'ajout de ces médecins dans la base des médecins actifs) des médecins reprenant une activité (ainsi que la sélection des médecins qui décèdent ou qui partent à la retraite parmi les médecins en cessation temporaire d'activité), par exemple, pourrait être effectuée plus tôt.

Les calculs présentés ci-dessus sont réalisés **successivement pour chaque année** de la période de projection, pour une itération donnée. **Les calculs correspondant à un scénario sont réalisés successivement x fois, x étant le nombre d'itérations** choisi. Pour les projections de 2008, 5 itérations ont été réalisées pour chaque scénario.

#### Encadré 1 - Aperçu général d'une simulation

Le modèle de projections d'effectifs de médecins est composé de **quatre modules**, chacun d'entre eux regroupant un certain nombre d'événements modélisés (voir A. 1. c.).

Le **module 1** simule le **déroulement des études médicales**, c'est-à-dire :

- \* le déroulement des deux premiers cycles des études médicales, du PCEM 2 au DCEM 4 (sous-module 1.1)
- \* l'affectation des étudiants en première année de TCEM à l'issue des ECN (sous-module 1.2)
- \* le déroulement du troisième cycle des études médicales (sous-module 1.3).

Le **module 2** simule l'**entrée dans la vie active** des jeunes diplômés.

- \* Il sélectionne les jeunes diplômés qui exerceront la médecine en France et attribue un délai à l'entrée dans la vie active aux entrants dans la vie active (sous-module 2.1), puis
- \* un sexe et un âge (sous-module 2.2)
- \* une région d'exercice (sous-module 2.3)
- \* un mode d'exercice (sous-module 2.4)
- \* une zone d'exercice (sous-module 2.5).

Le **module 3** simule les **changements de situation** qui interviennent au cours de la vie active des médecins, c'est-à-dire :

- \* le changement de région d'exercice (sous-module 3.1)
- \* le changement de mode d'exercice (sous-module 3.2)
- \* le changement de zone d'exercice (sous-module 3.3)
- \* la cessation temporaire d'activité (sous-module 3.4)
- \* la reprise d'activité (sous-module 3.5).

Le **module 4**, enfin, simule les **sorties définitives de la vie active** des médecins :

- \* le départ à la retraite (sous-module 4.1)
- \* le décès (sous-module 4.2).

#### Zoom sur...

##### ❶ ... les sélections (ou tirages) simultanées

**Les médecins qui cessent temporairement leur activité, les médecins qui décèdent et les médecins qui partent à la retraite sont sélectionnés simultanément.** Parmi les médecins en cessation temporaire d'activité, ceux qui reprennent une activité, ceux qui décèdent et ceux qui partent à la retraite sont aussi sélectionnés en même temps. Enfin, les étudiants de troisième cycle qui abandonnent les études médicales et ceux qui redoublent sont sélectionnés simultanément.

Il s'agit en fait à chaque fois d'**appliquer les probabilités à un champ identique à celui sur lequel elles sont calculées.** Prenons l'exemple de la cessation temporaire d'activité, du départ à la retraite et du décès. Les probabilités de cesser temporairement son activité sont calculées sur le champ des médecins actifs au cours d'une année donnée, champ qui semble le plus « naturel ». Il en est de même des probabilités de partir à la retraite et des probabilités de décéder. Si l'on sélectionnait dans un premier temps les médecins qui cessent temporairement leur activité au cours de l'année, que l'on ôtait les enregistrements correspondant à ces médecins de la base des médecins actifs et que l'on sélectionnait ensuite les médecins qui partent à la retraite, ces derniers seraient sélectionnés parmi les médecins actifs moins ceux qui cessent temporairement leur activité au cours de l'année. Les

probabilités de partir à la retraite ne seraient donc pas appliquées à un champ identique à celui sur lequel elles ont été estimées. Pour cette raison, les médecins qui partent à la retraite, ainsi que les médecins qui décèdent, sont sélectionnés en même temps que les médecins qui cessent temporairement leur activité.

**Pour sélectionner simultanément  $c$  % de médecins qui cessent temporairement leur activité,  $d$  % de médecins qui décèdent et  $r$  % de médecins qui partent à la retraite**, on simule la réalisation d'une loi uniforme sur la population des médecins actifs, on sélectionne pour cesser temporairement leur activité ceux auxquels la valeur attribuée est inférieure à  $c/100$ , on sélectionne pour décéder ceux pour lesquels elle est comprise entre  $c/100$  et  $(c+d)/100$  et on fait partir à la retraite ceux pour lesquels elle est comprise entre  $(c+d)/100$  et  $(c+d+r)/100$ .

**Il aurait été possible de procéder autrement.** On aurait pu estimer, par exemple, les probabilités de cesser temporairement son activité sur le champ des médecins actifs, estimer les probabilités de décéder sur le champ des médecins actifs amputé des médecins qui cessent temporairement leur activité au cours de l'année<sup>3</sup>, et estimer les probabilités de partir à la retraite sur le champ des médecins actifs amputé des médecins qui cessent temporairement leur activité et des médecins qui décèdent au cours de l'année. Il aurait alors été possible de simuler successivement les cessations temporaires d'activité, les décès et les départs à la retraite sans créer d'incohérence entre le champ qui aurait servi à l'estimation des probabilités et le champ auquel elles auraient été appliquées.

#### ② ... le moment choisi pour l'attribution du sexe et de l'âge aux jeunes médecins

Il aurait été possible d'attribuer un sexe et un âge non pas aux entrants dans la vie active, mais aux étudiants en médecine, éventuellement dès leur entrée en PCEM 2.

Cependant, **il aurait alors fallu estimer des taux** d'abandons, de redoublements, d'arrivées, des taux de « fuites », etc. **par sexe et par âge**, ou à défaut, faire l'hypothèse que les comportements en la matière sont invariants selon le sexe et l'âge (ce que l'on fait, finalement, en n'attribuant pas de sexe et d'âge aux étudiants).

Or, les différentes probabilités (d'abandonner, de redoubler, etc.) déclinées par sexe et âge **auraient été estimées sur les sous-populations de petites tailles, et par conséquent probablement instables.** Elles auraient également été appliquées à des sous-populations de petites tailles, ce qui aurait rendu les résultats des tirages plus aléatoires.

Parce qu'on n'attribue pas de sexe et d'âge aux étudiants, il n'est pas possible de produire des effectifs projetés d'effectifs d'étudiants par sexe et âge.

#### ③ ... l'ordre choisi pour l'attribution des caractéristiques aux entrants dans la vie active ou aux médecins reprenant leur activité, et pour la simulation des changements au cours de la vie active

L'ordre choisi pour l'attribution des caractéristiques aux entrants (sexe, âge, région, mode et zone d'exercice) **comporte une part d'arbitraire.** Nous avons choisi d'attribuer successivement le sexe et l'âge, la région d'exercice, le mode d'exercice et enfin la zone d'exercice.

Cet ordre est en fait *a priori* anodin, **du fait de la possibilité de décliner les probabilités de vivre chaque événement selon les critères les plus discriminants.** Par exemple, si le choix du mode d'exercice est fortement dépendant de celui de la région d'exercice, on cherchera à tenir compte de

---

<sup>3</sup> Sous réserve que les données le permettent.

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 dépendant de celui de la zone d'exercice, attribuée après le mode d'exercice aux nouveaux entrants. 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, il n'est finalement pas possible de retenir tous les critères que l'on souhaiterait : les strates sur lesquelles sont estimées les probabilités et dans lesquelles sont réalisés les tirages seraient trop petites (voir encadré 3). **Parce que le choix des critères est contraint, l'ordre des événements modélisés n'est donc en réalité pas forcément sans conséquence sur les résultats.**

Le choix de l'ordre pour l'attribution d'une région, d'un mode et d'une zone d'exercice aux médecins reprenant une activité, ainsi que pour la simulation des changements de région, de mode et de zone d'exercice a été fait dans les mêmes conditions.

## ***5. Les résultats produits par le modèle***

Le résultat produit par un modèle de microsimulation est une table de données individuelles pour chaque année de la période de projection. Pour limiter le volume de données stockées, les **résultats** ont été stockés sous la forme d'effectifs, calculés à partir de ces tables de données individuelles générées par le modèle. Le modèle produit donc des effectifs projetés de médecins par sexe, âge, spécialité (et généraliste/spécialiste), région d'exercice, mode d'exercice et zone d'exercice, mais aussi par région de diplôme, année d'obtention du diplôme et année d'entrée dans la vie active (les effectifs sont calculés pour chaque strate correspondant au croisement de toutes ces variables). Ils sont produits pour chaque année de la période de projection et pour chaque racine ou itération (voir A. 1. d.).

Le modèle produit d'autres tables de résultats, permettant de compléter l'analyse. Ces résultats portent sur les effectifs d'étudiants en médecine, les diplômés délivrés, les « fuites » (voir C. 1. a.), les répartitions des postes ouverts, pourvus et non pourvus aux ECN, les entrées et sorties de la vie active, les cessations temporaires et reprises d'activité, les changements de région, de mode, de zone d'exercice (etc.), pour chaque année de la période de projection.

Le modèle produit également des **tables utilisées en entrée** par le modèle. Ces tables contiennent essentiellement les probabilités qui sont appliquées aux médecins au cours de la simulation, ainsi que les données individuelles sur les médecins actifs et sur les médecins en cessation temporaire au début de la période de projection.

Enfin, le modèle produit des **tables de contrôle**, au fur et à mesure de la création des tables en entrée et du déroulement d'une simulation. Ces tables permettent d'effectuer des vérifications et de suivre a posteriori le déroulement des calculs, tant pour l'estimation des probabilités que pour la projection elle-même.

## B – Module 1 - Les études médicales

La nécessité de prévoir la possibilité de réaliser différentes simulations portant par exemple sur une variation du niveau du *numerus clausus*, ou sur une variation de la répartition des postes ouverts aux épreuves classantes nationales (ECN), a rendu indispensable la modélisation du déroulement du cursus universitaire de médecine.

### *1. Sous-module 1.1 : Du PCEM 2 (deuxième et dernière année de premier cycle des études médicales) au DCEM 4 (quatrième et dernière année de deuxième cycle)*

#### *a. Le choix de la méthode*

Dans un premier temps, le modèle simule le déroulement des deux premiers cycles des études médicales. Le résultat attendu de ce premier sous-module est le nombre d'étudiants inscrits en DCEM 4 à l'issue d'une année donnée de la période de projection. Cet effectif constitue une entrée du sous-module 1.2, qui simule le déroulement des ECN, épreuves auxquelles doivent être inscrits de manière systématique tous les étudiants de DCEM 4.

Toutefois, une contrainte supplémentaire a été ajoutée : ce premier sous-module doit donner également la répartition des étudiants de DCEM 4 par région (afin de permettre à l'avenir la modélisation d'une répartition régionale des postes ouverts aux ECN qui deviendrait dépendante du nombre d'étudiants en dernière année de DCEM dans chaque région).

Pour aboutir au nombre d'inscrits en DCEM 4 par région, la **modélisation agrégée** a été employée. Elle suppose la construction de quelques formules assez simples. C'est cette méthode qui était utilisée dans la version antérieure du modèle.

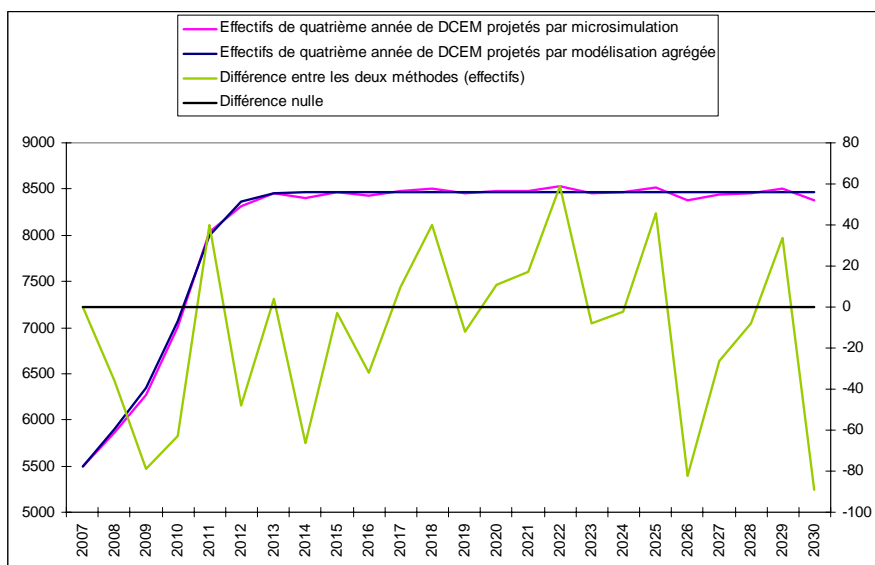
Toutefois, il aurait été possible d'utiliser la microsimulation pour ce sous-module. Les deux méthodes ont d'ailleurs été testées et leurs résultats comparés.

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 (voir graphique 1<sup>4</sup>). 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.

---

<sup>4</sup> La simulation a été réalisée à partir de données-test. Les valeurs du graphique sont donc fictives.

Graphique 1



Les résultats de la microsimulation sur cette partie du modèle ont également fait l'objet d'une étude approfondie, qui a montré que la microsimulation ne permet pas de générer, pour chaque année de la période de projection et chaque région, un taux de réalisation « d'évènements » (d'abandons, de redoublements, etc.) correspondant exactement au taux utilisé en entrée du modèle. Ceci s'explique par le côté aléatoire du « tirage des individus vivant les évènements », mais aussi et surtout par les effectifs modestes correspondant à chaque niveau d'études pour une région donnée, dans lesquels sont sélectionnés les individus vivant les évènements. Toutefois, en moyenne, sur toute la période de projection, au niveau régional, les taux obtenus avec la microsimulation sont très proches des taux utilisés en entrée du modèle.

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<sup>4</sup> (tableau 2) :

Tableau 2

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	3,3

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

Même si la microsimulation n'est pas utilisée pour cette partie du modèle, ces résultats sont importants, la microsimulation étant utilisée pour d'autres parties du modèle.

La méthode employée dans le cadre des projections d'effectifs de dentistes et de pharmaciens pour déterminer le nombre de diplômés, que nous appelons ici « méthode du ratio », consiste à appliquer aux effectifs d'entrants dans le cursus (après la barrière du *numerus clausus*) le ratio observé au cours des années récentes entre le nombre d'entrants et le nombre de diplômés. Cette méthode aurait également pu être employée ici, non pas pour déterminer directement le nombre de diplômés, puisque la modélisation des étapes intermédiaires (en particulier des ECN) est nécessaire à la simulation de certains scénarios, mais pour déterminer, à partir du nombre d'entrants en PCEM 2, le nombre d'étudiants en DCEM 4 cinq ans plus tard.

Cette solution a été écartée pour deux raisons principales. D'une part, la modélisation des étapes intermédiaires (passages d'un niveau à l'autre, abandons, redoublements, etc.) ouvre des perspectives supplémentaires : on peut ainsi simuler l'effet d'un changement des taux de redoublements ou d'abandons par exemple. D'autre part, modéliser chacun des événements permet d'observer en détail ce que simule le modèle, c'est-à-dire de voir combien d'abandons, de redoublements, de passages sont générés, ce qui permet de mieux analyser par la suite les résultats des projections.

Toutefois, cette méthode du ratio a été mise en œuvre et ses résultats ont été comparés à ceux produits par la modélisation agrégée. Ils sont proches : l'écart est de l'ordre de 3 %.

Les différents essais réalisés ont montré que ces méthodes ne donnent pas les mêmes résultats lorsque les variations du *numerus clausus* sont fortes ou lorsque les taux (de redoublements, d'abandons, etc.) sont élevés.

Dans la mesure où les données utilisées pour calculer les taux (de redoublements, d'abandons, etc.), servant à la modélisation du cursus médical semblent fiables, nous n'avons pas vu d'obstacle à l'utilisation de celle-ci. Si les taux obtenus avaient été jugés de qualité insuffisante, il aurait fallu recourir à la « méthode du ratio ».

#### ***b. La simulation : du numerus clausus aux effectifs en DCEM 4***

Les données utilisées en entrée de ce sous-module sont les données de SISE sur les inscriptions, sauf pour ce qui concerne le calcul des taux d'abandons et de redoublements du DCEM 4 (voir B. 2. c.).

Il n'est pas tenu compte du nombre d'étudiants en PCEM 1, dans la mesure où il existe une barrière à l'entrée en PCEM 2 : le *numerus clausus*, dont le niveau est fixé pour chaque région par les pouvoirs publics chaque année. La modélisation commence donc à la deuxième année de PCEM.

Les calculs exposés ci-après sont réalisés pour chaque année de la période de projection région par région, pour obtenir les effectifs régionaux d'étudiants inscrits en DCEM 4 à l'issue d'une année de la période de projection, que l'on somme pour obtenir l'effectif national.

Une hypothèse est faite sur le niveau national et sur la répartition régionale du *numerus clausus* de chacune des années de la période de projection, qui sont ainsi fixés ex-ante. Cette hypothèse peut varier selon le scénario retenu. On peut par exemple faire l'hypothèse d'un *numerus clausus* dont le niveau national et la répartition par région sont identiques aux derniers observés (observés au cours de

l'année précédant le début de la période de projection), et maintenus constants tout au long de la période de projection (voir B. 1. d.).

Les taux moyens de dépassement, de redoublements, d'abandons et d'arrivées estimés à partir des taux observés au cours des dernières années (voir B. 1. c.) sont appliqués à chacune des années de la période de projection. L'hypothèse sous-jacente par défaut est donc celle de comportements constants en ce qui concerne les dépassements du *numerus clausus*, les redoublements, les abandons et les arrivées (voir A. 3.).

Pour l'année courante de la période de projection, on part, pour une région donnée, de son *numerus clausus* et de son taux de dépassement (que nous calculons à partir du nombre d'inscrits, duquel on soustrait les redoublants pour ne tenir compte que des entrants) pour calculer le nombre d'entrants en deuxième année de PCEM. S'ajoutent aux entrants en deuxième année de PCEM les redoublants. Notons que le taux de dépassement du *numerus clausus* inclut les arrivées en PCEM 2. Les abandons viennent diminuer l'effectif. L'effectif d'inscrits en première année de DCEM correspond à l'effectif d'inscrits en deuxième année de PCEM l'année précédente, duquel il faut ôter les étudiants ayant abandonné ou redoublé à l'issue de l'année précédente et ajouter le nombre d'inscrits en première année de DCEM de l'année précédente qui redoublent, ainsi que les arrivants. Les effectifs sont calculés de la même manière pour les autres niveaux d'études.

Pour initialiser le calcul, on utilise les effectifs d'inscrits dans chacun des niveaux d'études au cours de la dernière année observée (2006-2007 pour les projections de 2008).

On fait l'hypothèse qu'il n'y a pas de mobilité interrégionale au cours des deux premiers cycles des études médicales. Cette hypothèse est basée sur l'observation des taux de mobilité, calculés à partir des données de SISE, qui ont semblé trop faibles pour que la mobilité interrégionale soit modélisée, d'autant plus que les niveaux des flux interrégionaux n'ont aucune importance pour le modèle s'ils se compensent : seuls les soldes régionaux ont un sens, le modèle visant à produire uniquement des effectifs régionaux d'étudiants et non à suivre les trajectoires individuelles d'une part et les étudiants étant indifférenciés (selon leur sexe, leur âge par exemple) d'autre part.

Voici les formules correspondantes, ainsi que trois schémas, le premier représentant les événements modélisés (schéma 2), le deuxième (schéma 3) l'enchaînement des calculs et le dernier (schéma 4) le résultat des calculs.

### Les formules appliquées :

Pour une région donnée :

- $PCEM2 = \text{part de la région dans le NC national} * \text{NC national} * (1 + \text{taux de dépassement}) + PCEM2_{n-1} * \text{taux de redoublements en PCEM2}$   
**Pour calculer l'effectif en PCEM 2, on n'utilise pas de taux d'arrivées. En effet, les arrivées sont déjà incluses dans le taux de dépassement.**
- $DCEM1 = (PCEM2_{n-1} * (1 - \text{taux de redoublements en PCEM2} - \text{taux d'abandons en PCEM2}) + DCEM1_{n-1} * \text{taux de redoublements en DCEM1}) / (1 - \text{taux d'arrivées en DCEM1})$
- $DCEM2 = (DCEM1_{n-1} * (1 - \text{taux de redoublements en DCEM1} - \text{taux d'abandons en DCEM1}) + DCEM2_{n-1} * \text{taux de redoublements en DCEM2}) / (1 - \text{taux d'arrivées en DCEM2})$
- $DCEM3 = (DCEM2_{n-1} * (1 - \text{taux de redoublements en DCEM2} - \text{taux d'abandons en DCEM2}) + DCEM3_{n-1} * \text{taux de redoublements en DCEM3}) / (1 - \text{taux d'arrivées en DCEM3})$

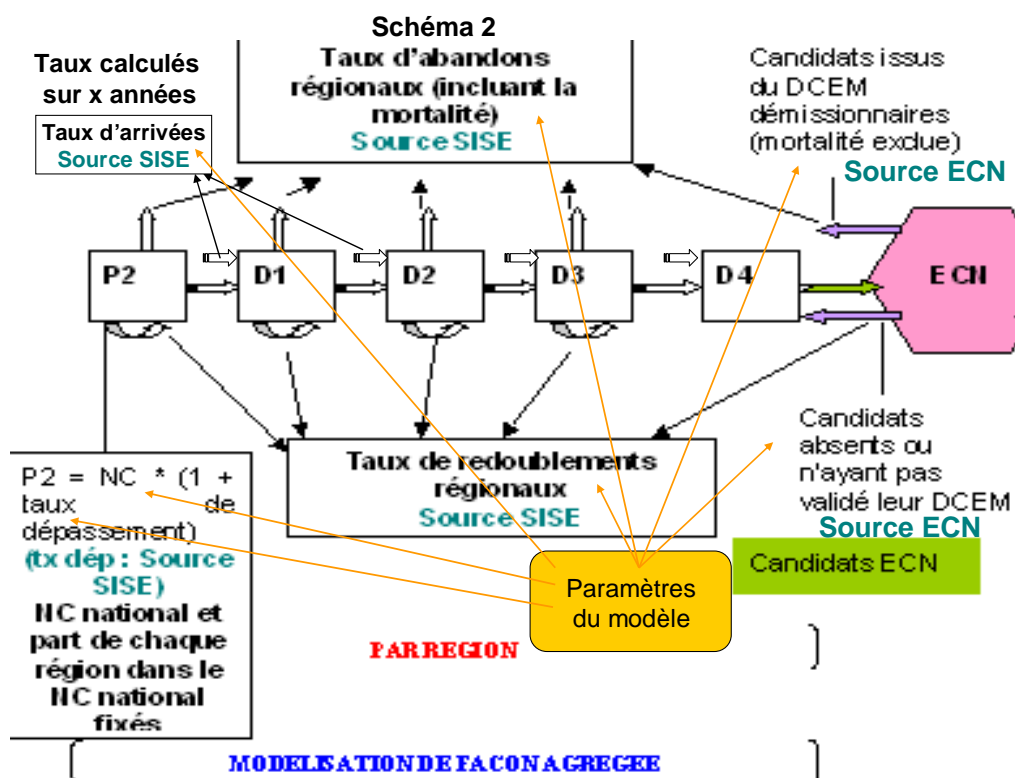
- $DCEM4 = (DCEM3_{n-1} * (1 - \text{taux de redoublements en DCEM3} - \text{taux d'abandons en DCEM3}) + DCEM4_{n-1} * \text{taux de redoublements en DCEM4}) / (1 - \text{taux d'arrivées en DCEM4})$

Pour le DCEM 4, on ne peut utiliser, comme pour les autres niveaux, le taux de redoublements calculé à partir des données de SISE. En effet, à l'issue des ECN, certains étudiants ne sont pas affectés parce qu'ils ne remplissent pas l'une des deux conditions nécessaires : ils n'ont pas validé leur DCEM ou ils ne se sont pas présentés aux ECN. Ils redoubleront donc leur dernière année de DCEM. Or, les redoublements que génère le sous-module 1.2 du modèle (qui simule l'affectation des étudiants à l'issue des ECN) à partir des taux d'absences aux ECN et d'invalidations du DCEM des candidats issus du DCEM 4 pourraient ne pas correspondre à ceux qui résulteraient des taux de redoublements issus de SISE. Pour garantir la cohérence entre les deux sous-modules concernés, les taux de redoublements appliqués aux étudiants en DCEM 4 dans le sous-module 1.1 sont donc ceux calculés à partir des données des ECN, qui sont également utilisés dans le sous-module 1.2.

Le taux d'abandons à l'issue du DCEM 4, qui pose le même problème que le taux de redoublements en DCEM 4, n'est quant à lui pas utile à la simulation (il n'est pas utilisé dans les formules).

On fait donc l'hypothèse que le nombre d'abandons et la mortalité (prise en compte dans le taux d'abandons) sont nuls en dernière année de DCEM.

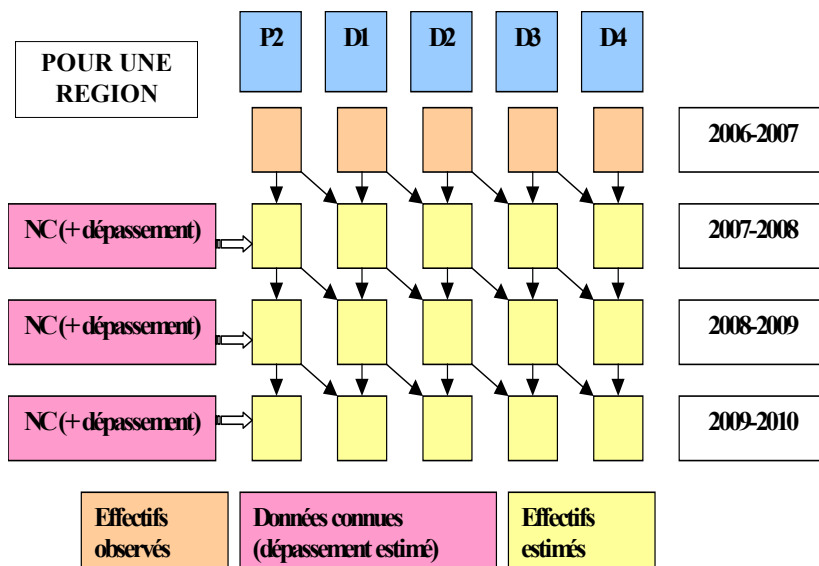
### Les événements modélisés





L'enchaînement des calculs

Schéma 3



La sortie du sous-module 1.1

Schéma 4

Année	Région	P2	D1	D2	D3	D4
2007-2008	Rég 1	Effectif	Effectif	Effectif	Effectif	Effectif
2007-2008	Rég 2	Effectif	Effectif	Effectif	Effectif	Effectif
...						
2007-2008	Rég 22	Effectif	Effectif	Effectif	Effectif	Effectif
2008-2009	Rég 1	Effectif	Effectif	Effectif	Effectif	Effectif
2008-2009	Rég 2	Effectif	Effectif	Effectif	Effectif	Effectif
...						
2008-2009	Rég 22	Effectif	Effectif	Effectif	Effectif	Effectif

*c. À propos des taux*

- Les taux d'abandons, de redoublements, d'arrivées et de dépassement du *numerus clausus* sont calculés, comme indiqué plus haut, à partir des données de SISE relatives aux inscriptions.
- Les différents taux calculés et appliqués aux effectifs projetés sont des **taux régionaux**. L'objectif est en effet de produire des effectifs régionaux, qui sont déterminés par des comportements parfois

nettement différenciés d'une région à l'autre en matière de dépassement du *numerus clausus*, d'abandons, de redoublements et d'arrivées, si l'on en croit justement les taux estimés.

• Concernant le mode de calcul des différents taux, précisions que :

- Les abandons pour le niveau d'études n (c'est-à-dire entre le niveau d'études n et le niveau n+1) correspondent aux étudiants inscrits à un niveau n du PCEM ou du DCEM (compris entre le PCEM 2 et le DCEM 4) au cours d'une année scolaire a qu'on ne retrouve pas inscrits au cursus médical (tous niveaux confondus) dans le fichier SISE de l'année scolaire a+1. Le taux (régional) d'abandons pour le niveau d'études n est le rapport entre ces abandons pour le niveau n et pour une région r et les effectifs d'inscrits à ce niveau n dans la région r au cours de l'année scolaire a. Par définition, les taux d'abandons incluent donc la mortalité. Le taux d'abandons à l'issue du DCEM 4 n'est pas calculé car il n'est pas utilisé pour la simulation.
- Les redoublements pour le niveau d'études n correspondent aux étudiants inscrits à un niveau n du PCEM ou du DCEM (compris entre le PCEM 2 et le DCEM 4) au cours d'une année scolaire a qu'on retrouve inscrits à la préparation du même diplôme, dans le même niveau d'études, dans le fichier SISE de l'année scolaire a+1. Le taux (régional) de redoublements pour le niveau d'études n est le rapport entre ces redoublements pour le niveau n et pour une région r et les effectifs d'inscrits à ce niveau n dans la région r au cours de l'année scolaire a.

- Les arrivées pour le niveau d'études n correspondent aux étudiants inscrits à un niveau n du DCEM (pas d'arrivées pour le PCEM 2 : elles sont incluses dans le taux de dépassement) au cours d'une année scolaire a qu'on ne retrouve pas inscrits à la préparation du diplôme d'Etat de docteur en médecine (en PCEM ou DCEM, tous niveaux confondus) dans le fichier SISE de l'année scolaire a-1. Le taux (régional) d'arrivées pour le niveau d'études n est le rapport entre ces arrivées pour le niveau n et pour une région r et les effectifs d'inscrits à ce niveau n dans la région r au cours de l'année scolaire a.

Dans la formule utilisée pour calculer les effectifs d'inscrits en DCEM 1 par exemple, le terme  $[PCEM2_{n-1} * (1 - \text{taux de redoublements en PCEM 2} - \text{taux d'abandons en PCEM 2}) + DCEM1_{n-1} * \text{taux de redoublements en DCEM 1}]$  correspond à l'effectif d'inscrits en DCEM 1 hors arrivées, qu'on nomme  $\bar{A}$ . Si on note T l'effectif d'inscrits y compris les arrivées, A le nombre d'arrivées et  $\hat{a}$  le taux d'arrivées, nous avons :

$$A + \bar{A} = T$$

$$\text{soit } \hat{a}T + \bar{A} = T$$

$$\text{soit encore } T(1 - \hat{a}) = \bar{A}$$

$$\text{et finalement } T = \bar{A} / (1 - \hat{a})$$

ce qui nous amène à la formule utilisée :

$$DCEM1 = (PCEM2_{n-1} * (1 - \text{taux de redoublements en PCEM 2} - \text{taux d'abandons en PCEM 2}) + DCEM1_{n-1} * \text{taux de redoublements en DCEM 1}) / (1 - \text{taux d'arrivées en DCEM 1})$$

Les taux d'arrivées incluent les entrées d'étrangers intégrant le cursus médical en cours de DCEM. Les entrées d'étrangers en PCEM 2 sont quant à elles incluses dans le taux de dépassement.

Les réimmatriculations sont également incluses à la fois dans les taux d'abandons et dans les taux d'arrivées. Il y a donc compensation et annulation de l'effet des réimmatriculations (voir A. 2. b.).

Les arrivées incluent également les reprises d'études, les interruptions d'études étant quant à elles incluses dans les taux d'abandons.

- Le dépassement du *numerus clausus* pour une région r et une année a correspond à la différence entre l'effectif d'étudiants entrant en PCEM 2 au cours de l'année scolaire a (c'est-à-dire l'effectif d'inscrits en PCEM 2 dans la région r au cours de l'année scolaire a, duquel on ôte l'effectif de redoublants en PCEM 2) et le *numerus clausus* de la région r pour l'année a. Le taux de dépassement est donc le dépassement rapporté au *numerus clausus*, moins un.

- Dans les données de SISE utilisées pour estimer les différents taux, pour un certain nombre d'inscrits en niveau n du PCEM ou du DCEM dans la région r au cours d'une année scolaire a, l'évolution entre l'année a et l'année a+1 est « anormale » : par exemple, des étudiants sont inscrits en niveau n en a et en niveau n+2 en a+1. Ces anomalies sont dues à des erreurs de codage, et sont assez rares pour que l'on puisse faire l'hypothèse que les taux de passages peuvent être déduits des taux d'abandons et de redoublements (voir B. 1. b.). Dans le modèle, les étudiants pour lesquels la simulation n'aura pas généré d'abandon ou de redoublement à l'issue de l'année a passeront donc au niveau supérieur.

- Les taux estimés à partir des fichiers SISE qui sont appliqués aux effectifs projetés correspondent aux taux moyens observés au cours des dernières années, le nombre d'années pouvant être compris entre 1 et 6.

#### **Encadré 2 - Calcul des taux moyens**

Par « taux moyens », il faut comprendre « taux obtenus en calculant les moyennes de taux annuels ». L'autre option possible était de calculer les taux en sommant les événements (par exemple les abandons) survenus au cours des X dernières années et en rapportant cette somme d'événements à la somme des effectifs inscrits au cours des X dernières années. Ces deux méthodes ne sont pas équivalentes, au sens où elles ne produisent pas les mêmes résultats, même si ceux-ci sont en général très proches. La seconde est légèrement plus sensible à une forte volatilité des taux et nettement plus sensible à une forte volatilité des effectifs totaux sur lesquels sont calculés les taux. Si un taux de redoublements est particulièrement bas pour une année donnée, le taux moyen de redoublements calculé à l'aide de la deuxième méthode sera très légèrement « tiré vers le bas ». Si les effectifs d'inscrits pour l'une des années sont particulièrement élevés, à proportions de redoublants identiques, le taux de redoublements sera « tiré vers le haut » avec la deuxième méthode. Pour cette raison, la première méthode a été préférée à la seconde, et a été employée pour tous les taux utilisés par le modèle, tous modules confondus.

Au vu des valeurs des différents taux sur les six dernières années et étant donné que les ECN, mises en place en 2004, ont pu engendrer des modifications des comportements, il était préférable d'utiliser un taux moyen sur les trois dernières années pour les projections de 2008, au moins pour la modélisation des études médicales.

- Les taux estimés à partir des fichiers SISE ont été jugés « réalistes » par les membres du comité technique connaissant le déroulement et la réalité des études médicales au niveau régional. En outre, il a semblé au comité technique qu'il était légitime de tenir compte des abandons, malgré leur rareté.

- Quelques corrections ont cependant été effectuées sur certains taux visiblement erronés.

- Les taux utilisés par défaut en entrée par le modèle sont les taux moyens calculés à partir des données relatives aux dernières années observées, dont le calcul par le modèle est « automatique ». L'hypothèse sous-jacente à l'utilisation de ces taux est donc celle de comportements constants en matière de dépassement du *numerus clausus*, d'abandons, de redoublements et d'arrivées. Cependant, il est possible d'utiliser d'autres taux, pour tester d'autres hypothèses (voir B. 1. d.).

#### ***d. Les simulations possibles***

Il est possible de simuler les effets d'une variation du *numerus clausus* national, sous l'hypothèse d'une répartition régionale de ce *numerus clausus* constante, et de simuler les effets d'une variation de la répartition régionale du *numerus clausus* sous l'hypothèse d'un *numerus clausus* national constant<sup>5</sup>. Il est également possible de simuler simultanément une variation du *numerus national* et une variation de sa répartition régionale.

<sup>5</sup> Toutefois, du fait des hypothèses faites, et en particulier de celle portant sur la répartition régionale des postes ouverts et des postes pourvus aux ECN, supposées indépendantes de la répartition régionale du *numerus clausus*, une variation de la répartition du *numerus clausus* ne peut avoir que des effets très faibles, voir nuls.

En outre, il est possible de simuler l'effet d'un changement de comportements en matière de dépassement du *numerus clausus*, d'abandons, de redoublements et d'arrivées. Pour ce faire, il faut définir, pour chaque région et pour chaque niveau d'études, les taux correspondants (qui restent fixés pour toute la période de projection), qui remplaceront les taux de dépassement du *numerus clausus*, d'abandons, de redoublements et d'arrivées correspondant à l'hypothèse de comportements constants, calculés automatiquement par le modèle (à partir des données sur les années récentes).

Les paramètres modifiables pour simuler les différents scénarios sont donc pour le sous-module 1.1 :

- Le *numerus clausus* national (qui peut varier pour chaque année de la période de projection).
- La répartition régionale du *numerus clausus* national (qui peut varier pour chaque année de la période de projection).
- Les taux de dépassement régionaux (fixés pour toute la période de projection).
- Les taux d'abandons régionaux (fixés pour toute la période de projection).
- Les taux de redoublements régionaux (fixés pour toute la période de projection).
- Les taux d'arrivées régionaux (fixés pour toute la période de projection).

Concernant ces quatre derniers types de taux, sont paramétrables le nombre d'années récentes sur lesquelles sont calculés les taux correspondant à l'hypothèse de comportements constants ainsi que la valeur de ces taux (qui peut être modifiée pour tester d'autres hypothèses que celle de comportements constants).

## ***2. Sous-module 1.2 : du DCEM 4 au TCEM 1 (première année de troisième cycle des études médicales) ou l'affectation des étudiants en troisième cycle à l'issue des ECN***

### ***a. Le choix de la méthode***

À l'issue des calculs correspondant au sous-module 1.1, le modèle a produit les effectifs régionaux d'étudiants inscrits en DCEM 4 au cours d'une année de la période de projection. Ces effectifs seront l'une des deux entrées du sous-module 1.2 pour l'année suivante de la période de projection. Celui-ci simule les résultats des affectations en première année de troisième cycle des étudiants s'étant présentés aux ECN et ayant validé leur DCEM. La seconde entrée de ce module est l'effectif d'internes de première année (étudiants inscrits en première année de troisième cycle) repassant les ECN en vue d'obtenir une nouvelle affectation. Cet effectif est produit au cours du passage précédent du programme (pour l'année précédente de la période de projection), par le sous-module 1.3, qui simule le déroulement du troisième cycle des études médicales.

Le sous-module 1.2 doit produire, pour une année donnée de la période de projection, les effectifs d'étudiants affectés en première année de TCEM par discipline (on parle ici des disciplines ECN, qui sont au nombre de 11) et par région.

La modélisation agrégée s'impose donc ici naturellement, les entrées et sorties de ce sous-module étant des agrégats.

### ***b. La simulation : des candidats aux ECN aux affectations en TCEM 1***

Les données utilisées en entrée de ce sous-module sont les données du CNG sur les ECN (voir A. 2. b.), ainsi que les effectifs régionaux d'inscrits en DCEM 4 et l'effectif d'internes de première année repassant les ECN, produits respectivement par les sous-modules 1.1 et 1.3.

Les résidents en médecine générale candidats aux ECN, qui terminent leur troisième cycle, n'ont pas à être pris en compte pour la simulation des ECN. Ceux-ci sont voués à disparaître, et plus aucun d'entre

eux ne peut entrer en troisième cycle. Ils sont déjà inclus dans les effectifs inscrits en troisième cycle au début de la période de projection.

Pour modéliser les ECN, une solution simple consisterait à observer la répartition des postes pourvus pour les années passées et, pour chaque année projetée, à appliquer cette répartition au nombre de postes pourvus à l'issue des ECN, que l'on calculerait.

Cependant, on ne peut alors simuler ni les effets d'une variation du nombre total de postes ouverts, ni ceux de la variation de la répartition du nombre total de postes ouverts entre la médecine générale et les autres disciplines ECN par exemple. Or, il s'agit là de variantes qu'il faut pouvoir simuler. De même, il est souhaitable de permettre la simulation des effets d'une variation de la répartition par discipline des postes ouverts dans les disciplines autres que la médecine générale, ou bien encore les effets d'une variation de la répartition régionale des postes ouverts aux ECN.

Afin de permettre toutes ces simulations, le sous-module 1.2 enchaîne les étapes suivantes :

①

- Calcul de la répartition actuelle (moyenne – sur les dernières années) des postes ouverts en médecine générale par région et de la répartition actuelle des postes ouverts dans les autres disciplines par région et par discipline (option « ECN n°1 », utilisée pour effectuer une simulation basée sur l'hypothèse d'une répartition par discipline des postes ouverts aux ECN dans les autres disciplines que la médecine générale constante) ou de la répartition actuelle des postes ouverts dans chacune des disciplines autres que la médecine générale par région (option « ECN n°2 », utilisée pour simuler les effets d'une variation de la répartition par discipline des postes ouverts dans les disciplines autres que la médecine générale).<sup>6</sup>

- Calcul de la répartition actuelle (moyenne – sur les dernières années) des postes non pourvus par région et par discipline.

Ces calculs sont réalisés une seule fois, avant la simulation des affectations de la première année de la période de projection.

②

Pour chaque année de la période de projection :

- Calcul de C, le nombre de candidats aux ECN comme suit :

**$C = \text{inscrits en dernière année de DCEM (sortie du sous-module 1.1) + internes de première année repassant les ECN (déduits de la sortie du sous-module 1.3)}$** <sup>7</sup>

Pour calculer C, le modèle a besoin d'estimer au préalable la proportion d'internes de première année repassant les ECN (voir B. 2. c.).

- Calcul du nombre total de postes ouverts (O) comme suit :

**$O = \alpha C$**  où C est le nombre de candidats aux ECN et  $\alpha$  un paramètre indiquant la « politique » retenue (nombre de postes ouverts inférieur ou supérieur de  $\alpha$  % au nombre de candidats).

---

<sup>6</sup> Distinguer ces deux cas n'a d'autre but que d'éviter, dans le cas où l'on souhaite faire l'hypothèse d'une répartition par discipline constante des postes ouverts dans les disciplines autres que la médecine générale, d'éviter de devoir calculer et saisir dans un fichier en entrée cette répartition. L'option 2 est équivalente à l'option 1 si l'on suppose constante la part de chaque discipline parmi les postes ouverts dans les disciplines autres que la médecine générale.

<sup>7</sup> Pour calculer le nombre de candidats aux ECN d'une année donnée de la période de projection, le modèle utilise le nombre d'inscrits en dernière année de DCEM et le nombre d'internes de première année estimés pour l'année précédente.

- Calcul du nombre total de postes ouverts en médecine générale (OMG) et du nombre total de postes ouverts dans les autres disciplines (OAU) comme suit :

**OMG =  $\beta$  O** où  $\beta$  est la proportion de postes ouverts en médecine générale par rapport au nombre total de postes ouverts

**OAU = O – OMG** ou **OAU = (1 -  $\beta$ ) O**

- Pour l'option « ECN n°2 », calcul du nombre de postes ouverts dans chacune des disciplines autres que la médecine générale :

**ODIS<sub>i</sub> = d<sub>i</sub>OAU** où d<sub>i</sub> est la part de la discipline i dans l'ensemble des postes ouverts dans les disciplines autres que la médecine générale, avec :

**ODIS1 + ODIS2 + ... + ODIS10 = OAU**

- Calcul du nombre total de postes pourvus (P) comme suit :

**P = C – candidats non affectés**

**P = C – candidats inscrits en dernière année de DCEM absents aux ECN ou n'ayant pas validé leur DCEM – candidats inscrits en dernière année de DCEM démissionnaires – candidats internes non affectés**

Pour calculer P, le modèle a besoin de calculer la proportion de candidats inscrits en dernière année de DCEM absents aux ECN ou n'ayant pas validé leur DCEM, la proportion de candidats inscrits en dernière année de DCEM démissionnaires, ainsi que la proportion d'internes non affectés parmi les candidats internes de première année (voir B. 2. c.).

- Calcul du nombre total de postes non pourvus (NP) comme suit :

**NP = O - P**

③

Pour chaque année de la période de projection :

- Appliquer la répartition observée des postes ouverts en médecine générale (calculée en ) au nombre total de postes ouverts en médecine générale (OMG) et la répartition observée des postes ouverts dans les autres disciplines au nombre total de postes ouverts dans les autres disciplines (OAU) pour obtenir la répartition des postes ouverts par région et par discipline de l'année.

- Appliquer la répartition observée des postes non pourvus (calculée en ) au nombre total de postes non pourvus (NP) pour obtenir la répartition des postes non pourvus par région et par discipline de l'année.

- Par soustraction des deux matrices ainsi créées, calculer la répartition des postes pourvus (donc des candidats affectés) par région et par discipline de l'année.

Le schéma 5 représente le calcul des matrices des postes ouverts et des postes non pourvus.

## Schéma 5

Pour l'option 1

Répartitions observées (source : ECN – CNG)

Postes ouverts					
	Rég 1	Rég 2	...	Rég n	Total
Dis 1					
Dis 2					
...					
Dis 10					
Total autres dis					100%
MG					100%

Postes non pourvus					
	Rég 1	Rég 2	...	Rég n	Total
Dis 1					
Dis 2					
...					
Dis 10					
MG					
Total					100%

Répartitions estimées par le modèle

Postes ouverts					
	Rég 1	Rég 2	...	Rég n	Total
Dis 1					ODIS1
Dis 2					ODIS2
...					
Dis 10					ODIS10
Total autres dis					OAU
MG					OMG

Postes non pourvus					
	Rég 1	Rég 2	...	Rég n	Total
Dis 1					
Dis 2					
...					
Dis 10					
MG					
Total					NP

Le schéma 6 représente la matrice des postes pourvus, c'est-à-dire la sortie du sous-module 1.2.

## Schéma 6

Répartition des postes pourvus (donc des candidats affectés) par région et par discipline

Postes pourvus					
	Rég 1	Rég 2	...	Rég n	Total
Dis 1					
Dis 2					
...					
Dis 10					
MG					
Total					P

### c. À propos des taux et des répartitions estimées

- Il est possible de prendre en compte un nombre variable d'années (de la seule dernière aux cinq dernières années) pour le calcul des répartitions observées de postes ouverts et de postes non pourvus et pour le calcul des taux d'absentéisme ou d'invalidations du DCEM des candidats issus du DCEM 4, des taux de démissions des candidats issus du DCEM 4, du taux d'internes de première année repassant les ECN et du taux de candidats internes non réaffectés.

Pour les projections de 2008, on ne pouvait cependant utiliser que les quatre dernières années, puisqu'au moment de la réalisation des projections, seulement 4 éditions des ECN avaient eu lieu (2004, 2005, 2006 et 2007).

- Pour calculer C, le nombre de candidats aux ECN, où

$C = \text{inscrits en dernière année de DCEM (sortie du sous-module 1.1)} + \text{internes de première année repassant les ECN (déduits de la sortie du sous-module 1.3)}$

le sous-module 1.2 utilise la sortie du sous-module 1.1 (les effectifs régionaux d'inscrits en DCEM 4, qu'il somme pour obtenir l'effectif national d'inscrits en dernière année de DCEM). Il utilise également la sortie du sous-module 1.3, donnant les effectifs par région, par discipline et par niveau des internes : il en déduit le nombre d'internes de première année pour l'année courante de la période de projection. Pour obtenir le nombre d'internes de première année repassant les ECN, il faut cependant multiplier ce nombre par la proportion nationale d'internes de première année repassant les ECN. Le modèle estime cette proportion en rapportant, pour chaque année des données ECN utilisées en entrée, l'effectif de candidats aux ECN internes de première année, issu des données relatives aux ECN du CNG, à l'effectif d'inscrits en première année de TCEM, issu des données SISE. Le taux utilisé en entrée est la moyenne des taux annuels. Cette moyenne est en principe calculée sur toutes les années des données relatives aux ECN utilisées. Toutefois, si la plus ancienne des années utilisées est 2004, cette moyenne ne peut être calculée que sur le nombre d'années de données relatives aux ECN utilisées moins une. En effet, avant 2007, les internes ne peuvent être repérés parmi les candidats aux ECN dans les fichiers de données relatives aux ECN du CNG que par l'appariement de deux fichiers annuels successifs. Les internes ne peuvent donc pas être repérés dans le fichier de 2004.

Pour calculer P, le nombre de postes pourvus, où

$$P = C - \text{candidats inscrits en dernière année de DCEM absents aux ECN ou n'ayant pas validé leur DCEM} - \text{candidats inscrits en dernière année de DCEM démissionnaires} - \text{candidats internes non affectés,}$$

le sous-module 1.2 multiplie le nombre d'internes de première année repassant les ECN (voir ci-dessus) par la proportion nationale de candidats internes non réaffectés parmi les internes de première année candidats, pour obtenir le nombre de candidats internes non affectés. Cette proportion est estimée par le modèle, qui rapporte, pour chaque année des données ECN utilisées en entrée, l'effectif de candidats internes de première année non affectés à l'effectif de candidats internes de première année, tous deux issus des données relatives aux ECN du CNG. Le taux utilisé en entrée est la moyenne des taux annuels. L'observation faite au paragraphe précédent pour les données relatives à l'année 2004 vaut également pour le calcul de cette proportion.

Pour calculer P, le sous-module 1.2 utilise également la sortie du sous-module 1.1 (les effectifs régionaux d'inscrits en DCEM 4). Il multiplie l'effectif d'inscrits en DCEM 4 de chaque région par la proportion régionale d'étudiants issus du DCEM absents aux ECN ou n'ayant pas validé leur DCEM pour obtenir le nombre régional de candidats inscrits en dernière année de DCEM absents aux ECN ou n'ayant pas validé leur DCEM. Il multiplie ce même effectif par la proportion régionale d'étudiants issus du DCEM démissionnaires pour obtenir le nombre régional de candidats inscrits en dernière année de DCEM démissionnaires. La première de ces deux proportions est obtenue en rapportant, pour chaque année des données ECN utilisées en entrée, l'effectif régional de candidats issus du DCEM ne s'étant pas classés aux ECN (absents) ou n'ayant pas validé leur DCEM à l'effectif régional de candidats issus du DCEM. La seconde est obtenue en rapportant, pour chaque année des données ECN utilisées en entrée, l'effectif régional de candidats issus du DCEM s'étant classés aux ECN et ayant validé leur DCEM mais n'ayant pas choisi d'affectation à l'effectif régional de candidats issus du DCEM. Ces deux proportions sont calculées à partir des données relatives aux ECN du CNG. Les proportions utilisées en entrée du modèle sont les moyennes des proportions annuelles. Ces moyennes sont calculées sur toutes les années des données relatives aux ECN utilisées.

En sommant les effectifs régionaux obtenus, le modèle calcule l'effectif national de candidats inscrits en dernière année de DCEM absents aux ECN ou n'ayant pas validé leur DCEM et l'effectif national de candidats inscrits en dernière année de DCEM démissionnaires.

Le calcul de la proportion d'étudiants issus du DCEM 4 absents aux ECN ou n'ayant pas validé leur DCEM au niveau régional est nécessaire dans la mesure où elle fera office de taux de redoublements



en DCEM 4, ceci afin d'assurer la cohérence entre les résultats des sous-modules 1.1 et 1.2 (voir B. 1. b.). Ce taux est estimé à partir des données relatives aux ECN du CNG, contrairement aux mêmes taux pour les autres niveaux du PCEM et du DCEM. Tous seront cependant appliqués aux effectifs régionaux d'inscrits issus des données SISE 2006-2007 (pour la première année de projection) puis aux effectifs projetés.

Il n'est pas envisageable de calculer pour le sous-module 1.2 une proportion nationale d'étudiants issus du DCEM absents aux ECN ou n'ayant pas validé leur DCEM et pour le sous-module 1.1 des proportions régionales. En effet, les proportions nationales, appliquées à des effectifs nationaux, ne donneraient pas les mêmes effectifs que les proportions régionales appliquées à des effectifs régionaux, ce qui créerait une discordance entre les résultats des deux sous-modules.

- Les internes de première année ayant repassé les ECN et réaffectés quittent le poste qu'ils occupaient pour prendre leur nouvelle affectation. Le sous-module 1.3 « injectera » le flux d'étudiants correspondant aux affectations prononcées à l'issue des ECN dans les effectifs de première année de troisième cycle. Il faut donc au préalable enlever les effectifs d'internes réaffectés des effectifs de troisième cycle, parmi lesquels ils sont déjà comptabilisés (dans la région et la discipline correspondant à leur ancienne affectation), afin de ne pas créer de doubles-comptes. Pour ce faire, le sous-module 1.2 a calculé, pour chaque année des données ECN utilisées en entrée, la répartition des candidats aux ECN internes de première année réaffectés par région et par discipline « d'origine » (parmi ceux pour lesquels la variable indiquant la discipline de l'année précédente est renseignée, ce qui est presque toujours le cas). La répartition utilisée en entrée est la moyenne des répartitions annuelles. Cette moyenne est en principe calculée sur toutes les années des données relatives aux ECN utilisées, sauf dans le cas où l'année 2004 est utilisée. En effet, l'observation faite au paragraphe ci-dessus pour les données relatives à l'année 2004 vaut également pour le calcul de cette répartition.

- Pour modéliser les ECN, nous avons estimé la répartition des postes non pourvus à partir de la répartition actuelle des postes non pourvus par région et par discipline. Par défaut, on fait donc l'hypothèse que cette répartition reste inchangée au cours de la période de projection, et qu'elle est identique à celle observée actuellement. Or, cette hypothèse est assez forte. En effet, supposons que le nombre de postes ouverts en médecine générale soit fortement augmenté, en particulier dans certaines régions, dans lesquelles les pouvoirs publics souhaitent favoriser les affectations en médecine générale, il est probable que dans les faits, les postes non pourvus dans cette discipline et ces régions soient proportionnellement plus nombreux, c'est-à-dire que la répartition des postes non pourvus soit modifiée. Pour la simulation de certains scénarios (ceux dans lesquels la répartition des postes ouverts aux ECN est modifiée), cette hypothèse est importante. Il faut en tenir compte pour interpréter les résultats des simulations correspondantes.

#### ***d. Les simulations possibles***

Les paramètres modifiables pour simuler les différents scénarios sont pour le sous-module 1.2 :

- la « politique » retenue, c'est-à-dire le rapport entre le nombre de candidats et le nombre de postes ouverts ;
- la répartition des postes ouverts entre la médecine générale et les autres disciplines ;
- la répartition par discipline des postes ouverts dans les disciplines autres que la médecine générale ;
- les taux régionaux d'absences aux ECN ou d'invalidations du DCEM des candidats issus du DCEM 4 (fixés pour toute la période de projection) ;
- les taux régionaux de candidats issus du DCEM 4 démissionnaires, c'est-à-dire n'ayant pas choisi de poste (fixés pour toute la période de projection) ;
- la proportion nationale d'internes de première année repassant les ECN (fixée pour toute la période de projection) ;

- la proportion nationale d'internes candidats non réaffectés à l'issue des ECN (fixée pour toute la période de projection).

La répartition des postes ouverts et la répartition des postes non pourvus par région et par discipline peuvent être également modifiées pour simuler les effets d'hypothèses différentes quant à ces répartitions.

### ***3. Sous-module 1.3 : du TCEM 1 (première année de troisième cycle des études médicales) à l'obtention du diplôme***

#### ***a. Le choix de la méthode***

Le sous-module 1.2 du modèle produit la répartition des postes pourvus à l'issue des ECN, c'est-à-dire des affectations prononcées, par région et par discipline. Il s'agit donc de résultats agrégés, d'effectifs, à partir desquels on pourrait, par des formules et des itérations, déduire les effectifs d'étudiants inscrits dans chacun des niveaux du troisième cycle, comme pour les deux premiers cycles, afin d'aboutir au nombre de diplômés.

Comme le sous-module 1.1 (premier et deuxième cycles), le sous-module 1.3 doit produire des effectifs d'étudiants par région et par niveau d'études, mais il doit aussi décliner ces effectifs régionaux par discipline. Dès lors, les formules à construire seraient plus nombreuses. Pour le troisième cycle, la modélisation de certains événements, en particulier la sélection des internes de première année ayant repassé les ECN et réaffectés à l'issue des ECN (à ôter des effectifs en formation dans les différentes régions et disciplines avant de les réaffecter après la simulation des ECN) compliquerait également les formules. Pour ces deux raisons, nous avons choisi d'utiliser la microsimulation pour modéliser le déroulement du troisième cycle des études médicales, méthode qui rend la simulation plus simple à réaliser en pratique.

La sortie produite par le sous-module 1.2 (répartition des postes pourvus à l'issue des ECN par région et par discipline) est donc « transformée » en données individuelles : si pour une région et une discipline données, le nombre d'affectations prononcées est de  $x$ ,  $x$  enregistrements individuels correspondant à cette région et à cette discipline sont générés par le modèle.

De même, les effectifs d'inscrits dans chaque région, chaque discipline et chaque niveau du troisième cycle au moment du début de la période de projection sont transformés en données individuelles.

Dès lors, on peut microsimuler le déroulement du troisième cycle des études médicales : la base de données contenant les enregistrements individuels des inscrits en troisième cycle au début de la période de projection évoluera chaque année pour représenter les inscrits en troisième cycle au cours de chaque année de la période de projection.

Le sous-module 1.3 doit produire le nombre de diplômés pour chaque spécialité (regroupement des spécialités en 23 modalités – voir A. 1. a. - effectué pour les projections) et pour chaque région. Le sous-module 2.1 utilisera la sortie du sous-module 1.3 : il fera entrer dans la vie active ceux des diplômés qui exerceront tôt ou tard la médecine.

#### ***b. La simulation : des affectations en TCEM 1 aux diplômés***

Le troisième cycle se compose de trois années d'études pour la médecine générale, de cinq années pour les spécialités chirurgicales (à l'exception de la stomatologie, en voie de disparition : 1 à 2 étudiants par an), la gynécologie-obstétrique et l'anesthésie réanimation, de quatre années pour la

biologie médicale, la pédiatrie, la gynécologie médicale, la santé publique, la médecine du travail et la psychiatrie. En ce qui concerne la discipline « spécialités médicales », le nombre d'années de formation est de quatre pour certaines spécialités et de cinq pour les autres.

Pour simuler le déroulement du troisième cycle des études médicales, le sous-module 1.3 utilise une base de données individuelles indiquant pour chaque étudiant sa région, sa discipline et son niveau d'études au moment du début de la période de projection (en 2006-2007 pour les projections de 2008). Cette base est produite par le modèle lui-même, directement à partir des données SISE sur les inscriptions prises au cours de l'année précédant le début de la période de projection<sup>8</sup>. Le modèle la fait évoluer au fur et à mesure de la simulation pour qu'elle représente toujours la répartition par discipline, par région et par niveau d'études des internes en formation au cours de l'année courante de la période de projection.

Pour faire évoluer cette base, il enchaîne, pour une année donnée de la période de projection, les étapes suivantes :

- Le sous-module 1.3 doit ôter de la base individuelle les internes de première année qui ont repassé les ECN au cours de l'année courante de la période de projection et qui sont réaffectés en première année de troisième cycle, afin d'éviter les doubles-comptes. Ces internes feront en effet partie des flux d'entrants en première année de TCEM, qui seront intégrés aux effectifs d'étudiants en troisième cycle par le sous-module 1.3.

Dans un premier temps, le **nombre d'internes à ôter est calculé** à partir du nombre d'internes de première année, de la proportion nationale d'internes de première année se présentant aux ECN et de la proportion des internes non réaffectés parmi les internes de première année se représentant aux ECN, ces deux proportions ayant été estimées à partir des données observées sur les dernières années par le sous-module 1.2. Le modèle **calcule ensuite le nombre d'internes à ôter dans chaque région et chaque discipline**, en fonction de la répartition des internes réaffectés par discipline et par région, également estimée à partir des données observées sur les dernières années par le sous-module 1.2. La sélection des internes à ôter de la base de données individuelles n'est pas effectuée immédiatement (voir ci-après).

- Pour faire évoluer les internes dans le TCEM, le sous-module 1.3 applique également, comme le sous-module 1.1, des taux de redoublements, des taux d'abandons et des taux d'arrivées, qui sont estimés par défaut par le modèle lui-même à partir des données de SISE relatives aux inscriptions prises pour les années récentes. L'hypothèse sous-jacente est donc celle de comportements constants en la matière. Pour le TCEM, les taux sont des taux nationaux, mais calculés pour chaque discipline ECN (voir B. 3. c.).

La **sélection des internes qui abandonnent ou qui redoublent leur première année de TCEM** est effectuée **en même temps que la sélection des internes qui repassent les ECN et qui sont réaffectés à l'issue des épreuves**, à ôter de la base de données individuelles, afin que les probabilités (d'abandonner, de redoubler, de repasser les ECN et d'être réaffecté) soient appliquées au même champ que celui sur lequel elles sont calculées (c'est-à-dire l'ensemble des étudiants de première

---

<sup>8</sup> Les données issues de SISE permettent de calculer des effectifs d'étudiants en formation par région, niveau d'études et discipline au cours de l'année précédant le début de la période de projection. Le modèle crée à partir de ces effectifs une base de données individuelles contenant autant d'enregistrements que nécessaire pour chaque strate définie par la région, la discipline et le niveau d'études.

année de TCEM). Cependant, tandis que pour la sélection des étudiants qui abandonnent ou qui redoublent, on utilise un tirage aléatoire tel que celui décrit au A. 1. d., on emploie une autre méthode<sup>9</sup> pour sélectionner les internes à ôter de la base de données individuelles, le tirage aléatoire étant difficile à réaliser pour cette sélection, du fait des très petits effectifs concernés. Toutefois, du fait des arrondis, le nombre total d'internes sélectionnés finalement est en général un peu inférieur à l'effectif cible. De ce fait, le nombre d'internes est en fin de compte très légèrement surestimé.

**En même temps** qu'il sélectionne les étudiants de troisième cycle qui abandonnent les études médicales ou qui redoublent et, pour la première année, les internes à ôter de la base individuelle, le sous-module 1.3 **sélectionne les internes qui obtiennent le diplôme d'Etat de docteur en médecine à l'issue de l'année**, pour les 3<sup>ème</sup>, 4<sup>ème</sup> et 5<sup>ème</sup> années d'internat. Les enregistrements correspondant aux nouveaux diplômés sont placés dans une base de données individuelles à laquelle le sous-module 2.1, qui sélectionne ceux des diplômés qui exerceront un jour la médecine et leur affecte un délai à l'entrée dans la vie active, fera appel.

Pour déterminer les effectifs de diplômés par région et par spécialité qu'il doit produire, le sous-module 1.3 fait l'hypothèse que sur le niveau terminal de chaque spécialité (troisième année pour la médecine générale, quatrième pour la pédiatrie, etc.), le taux d'abandons est nul. Il suppose alors que tous les étudiants inscrits dans une région donnée à une année terminale du TCEM qui ne redoublent pas sont diplômés. Pour l'année terminale de la formation, il sélectionne donc, pour une région et une discipline données les étudiants qui redoubleront au cours de l'année scolaire suivante, et considère comme diplômés tous les autres étudiants. Cette méthode a été retenue par défaut, les autres moyens de calculer les effectifs de diplômés n'ayant pas pu être mis en œuvre (voir B. 3. c.).

Le cas des spécialités médicales est un peu particulier puisque dans les faits, les étudiants sont diplômés soit à l'issue de leur quatrième année de TCEM, soit à l'issue de leur cinquième année, selon leur spécialité. Les mauvais codages du niveau d'études repérés dans les données SISE sur les inscriptions pour les niveaux 4 et 5 de la discipline rendent très incertaine l'estimation d'un taux de passage entre la quatrième année et la cinquième année de TCEM pour la discipline « spécialités médicales ». Or, sans ce taux de passage, on ne peut, dans la simulation, distinguer les étudiants à diplômer à l'issue de la quatrième année de ceux à diplômer à l'issue de la cinquième année. On a donc choisi de regrouper, pour cette discipline, les étudiants inscrits en quatrième année et ceux inscrits en cinquième année dans un niveau fictif unique, nommé niveau 4b. Les étudiants qui, dans la réalité, passent du niveau 4 au niveau 5 redoublent dans le modèle. Ce faisant, on « mélange » des étudiants appartenant à deux promotions différentes, ce qui est cependant sans conséquence sur le calcul effectué, qui vise seulement à déterminer les effectifs de diplômés à l'issue du TCEM. On obtient donc le nombre de diplômés de la discipline « spécialité médicales » pour la région  $r$  à l'issue de l'année  $a$  en déduisant du nombre d'inscrits en niveau 4b dans cette discipline et dans la région  $r$  le nombre de redoublants à l'issue de l'année  $a$ .

• Pour une année donnée de la période de projection, le sous-module 1.3 **ajoute** à la base de données individuelles, relatives aux internes en formation, **les individus dont le sous-module 1.2 a simulé l'affectation en première année de troisième cycle à l'issue des ECN**. Le sous-module 1.2 produit des effectifs d'affectés par discipline et par région. Le sous-module 1.3 convertit ces informations en

---

<sup>9</sup> Si dans chaque strate définie par la région et la discipline, on doit ôter  $x$  internes, les  $x$  derniers étudiants de chaque strate sont sélectionnés, le fichier étant préalablement trié dans un ordre aléatoire, déterminé par une variable appelée « aléa », qui correspond à la simulation de la réalisation d'une loi uniforme. Afin d'éviter que les mêmes individus soient sélectionnés pour être ôtés de la base et pour abandonner ou redoubler, les individus à ôter de la base sont sélectionnés parmi ceux pour lesquels la valeur de la variable aléa est très élevée, tandis qu'on sélectionne pour abandonner les individus pour lesquels la valeur prise par la variable aléa est inférieure à la probabilité d'abandonner et pour redoubler ceux pour lesquels la valeur prise par la variable aléa est comprise entre la probabilité d'abandonner et la probabilité de redoubler. Du fait des valeurs assez faibles des probabilités d'abandonner ou de redoubler, les internes de première année sélectionnés pour être ôtés de la base individuelle n'ont ainsi aucune chance d'être sélectionnés également pour abandonner ou redoubler.

données individuelles : il génère autant d'enregistrements correspondant à des individus entrant en première année du TCEM dans la discipline  $d$  et la région  $r$  qu'en donne le sous-module 1.2.

• De même, le sous-module 1.3 **ajoute** dans cette base de données individuelles, pour chaque niveau d'études, chaque discipline et chaque région, **des enregistrements correspondant aux étudiants qui « arrivent »**. Au préalable, le nombre d'arrivées pour chaque année d'études, chaque discipline et chaque région est calculé par le sous-module 1.3, qui applique les taux d'arrivées estimés à partir de données observées récentes aux effectifs obtenus pour l'année courante de la période de projection hors arrivées.

• Enfin, **pour les disciplines spécialités médicales » et « spécialités chirurgicales »**, le sous-module 1.3 doit encore **ventiler les diplômés de l'année courante de la période de projection par spécialité**, ces deux disciplines regroupant différentes spécialités. Pour ce faire, il utilise une répartition de diplômés de la discipline « spécialités médicales » (resp. « spécialités chirurgicales ») par spécialité, estimée au préalable à partir de données observées récentes, et un tirage aléatoire tel que celui décrit au A. 1. d.

L'ordre (dans la simulation) des événements se produisant au cours du TCEM dont le sous-module 1.3 simule la réalisation a son importance (voir A. 4.).

Comme pour les deux premiers cycles des études médicales et pour les mêmes raisons, la mobilité interrégionale est supposée nulle au cours du troisième cycle.

### *c. À propos des taux et des répartitions estimées*

• Les taux estimés d'abandons, de redoublements et d'arrivées en troisième cycle sont définis de la même façon que ceux estimés pour le premier et le deuxième cycles<sup>10</sup>. Ils sont également calculés à partir des données issues de SISE relatives aux inscriptions. En revanche, pour le troisième cycle, il s'agit de **taux nationaux** et non de taux régionaux, mais déclinés par discipline. Estimer ces taux pour chaque discipline au niveau régional conduirait à travailler sur des effectifs trop faibles.

Le taux d'abandons est supposé nul pour les années terminales de chaque discipline, afin qu'il soit possible de déduire le nombre de diplômés du nombre d'étudiants inscrits en année terminale et du nombre de redoublants (voir B. 3. b.). La mortalité, en particulier, incluse dans le taux d'abandons, est donc supposée nulle.

Du fait des réimmatriculations, les taux d'abandons et d'arrivées sont légèrement biaisés, mais le biais des taux d'arrivées vient compenser celui des taux d'abandons (voir A. 2. b.), et réciproquement. Cependant, le biais du taux d'arrivées en première année de TCEM n'est pas compensé par le biais du taux d'abandons en DCEM 4, qui n'est pas calculé à partir des données de SISE relatives aux inscriptions, et qui n'est donc pas entaché du biais dû aux réimmatriculations. Le taux d'arrivées en première année de TCEM est donc légèrement surestimé.

Le calcul d'un taux national d'arrivées, en particulier, est critiquable, puisqu'on suppose ainsi qu'il n'existe pas de différence entre les régions en matière d'attractivité.

---

<sup>10</sup> Les arrivées correspondent cependant pour le troisième cycle aux étudiants inscrits en troisième cycle au cours d'une année  $n+1$  qui n'étaient pas inscrits à la préparation d'un diplôme de médecine au cours de l'année  $n$ , tandis que pour les deux premiers cycles, les « arrivants » sont les étudiants inscrits en premier ou deuxième cycle au cours d'une année  $n+1$  qui n'étaient pas inscrits à la préparation du diplôme d'Etat de docteur en médecine au cours de l'année  $n$ . Cette différence est cependant infime.

- La répartition des diplômés des disciplines « spécialités médicales » et « spécialités chirurgicales » par spécialité est estimée à partir des données de SISE relatives aux diplômes. Les étudiants qui, d'après ces données, ont validé à la fois le DES et la thèse sont diplômés. C'est la répartition de ces seuls étudiants par spécialité qui est observée et qui est appliquée par le sous-module 1.3 pour attribuer aux diplômés de ces deux disciplines une spécialité.

- Quelques anomalies ont été repérées dans les différents taux estimés (notamment pour la discipline « spécialités chirurgicales »), dues notamment à des erreurs de codage. Quelques redressements des taux concernés ont donc été effectués.

De même, dans les données de SISE relatives aux inscriptions, le niveau d'études a parfois été redressé, lorsqu'il n'existait pas pour la discipline. Les étudiants dont la seule inscription était une inscription en thèse ont été d'office considérés comme inscrits en année terminale (en général, ces étudiants ont déjà validé leur DES et n'attendent plus que la validation de leur thèse pour être diplômés).

- Comme pour les deux premiers cycles, les taux estimés à partir des fichiers SISE qui sont appliqués aux effectifs projetés correspondent aux taux moyens observés au cours des dernières années, le nombre d'années pouvant être compris entre 1 et 6.

- Les taux utilisés par défaut en entrée par le modèle sont les taux moyens calculés à partir des données relatives aux dernières années observées, dont le calcul par le modèle est « automatique ». L'hypothèse sous-jacente à l'utilisation de ces taux est donc celle de comportements constants en matière d'abandons, de redoublements et d'arrivées. Cependant, il est possible d'utiliser d'autres taux, pour tester d'autres hypothèses.

- Le sous-module 1.3 estime le nombre de diplômés de la spécialité « gynécologie médicale », qui constitue une discipline des ECN. Cependant, pour les projections, les faibles effectifs de cette spécialité sont projetés avec ceux d'autres spécialités, qui composent la « spécialité » (au sens des projections) « Autres spécialités ». Par conséquent, pour les diplômés de gynécologie médicale, le sous-module 1.3 recode la spécialité.

- Les données de SISE relatives aux inscriptions n'indiquent pas, pour les étudiants inscrits uniquement à la préparation de la thèse, la spécialité préparée. Par appariement entre les différents fichiers annuels deux à deux, la spécialité de ces étudiants a été retrouvée lorsqu'ils étaient inscrits une année donnée à la préparation de la thèse uniquement, et l'année précédente au DES. On a fait l'hypothèse que leur spécialité n'avait pas changé entre les deux années, ce qui est extrêmement probable.

- Pour le calcul des effectifs d'étudiants inscrits en TCEM en biologie médicale avant le début de la période de projection, on a ôté des fichiers SISE des inscriptions un certain nombre d'enregistrements, selon une proportion qu'on a estimée (en comparant, pour quelques années données, le nombre d'inscrits en première année du DES de biologie médicale d'après les données de SISE au nombre de postes ouverts aux ECN dans cette discipline), afin de ne pas surestimer le nombre d'internes de cette discipline. En effet, les données de SISE ne distinguent pas les biologistes médecins des biologistes pharmaciens.

- Le mode de calcul retenu pour l'estimation des effectifs de nouveaux diplômés par région et par spécialité a été choisi par défaut.

On a cherché à estimer des taux de diplômés (desquels déduire directement les effectifs de diplômés), en rapportant les effectifs de diplômés aux effectifs d'inscrits (tous deux calculés à partir des

différentes sources disponibles), mais les estimations obtenues se sont avérées trop approximatives, du fait des défauts des sources (voir A. 2. b.).

Il aurait été possible également de supposer que dans les données issues de SISE, les étudiants inscrits une année donnée en année terminale et inscrits en DESC l'année suivante ou que l'on ne retrouve pas dans le fichier de l'année suivante étaient diplômés. Cependant, du fait de quelques erreurs de codage, les taux de diplômés ainsi obtenus n'étaient pas tous corrects. Les erreurs de codage pouvaient être corrigées, mais ces corrections auraient nécessité des calculs supplémentaires et induit une marge d'erreur au moins égale à celle qui est associée à la méthode finalement retenue.

#### *d. Les simulations possibles*

Il est possible de simuler l'effet d'un changement de comportements en matière d'abandons, de redoublements et d'arrivées. Pour ce faire, il faut définir, pour chaque discipline et pour chaque niveau d'études, les taux correspondants (qui restent fixés pour toute la période de projection). Le modèle calcule automatiquement les taux d'abandons, de redoublements et d'arrivées correspondant à l'hypothèse de comportements constants (à partir des données sur les années récentes). En modifiant les taux de redoublements appliqués aux années terminales, on peut également simuler les effets d'un changement dans les « taux de diplômés » (les étudiants ne redoublant pas sur une année terminale étant par défaut diplômés par le modèle).

Il est possible aussi de modifier la répartition des diplômés de la discipline « spécialités médicales » par spécialité, ainsi que la répartition des diplômés de la discipline « spécialités chirurgicales » par spécialité.

Concernant ces taux ou répartitions, sont paramétrables le nombre d'années récentes sur lesquelles sont calculés les taux ou répartitions correspondant à l'hypothèse de comportements constants ainsi que la valeur de ces taux ou répartitions (que l'on peut modifier pour faire d'autres hypothèses que celle de comportements constants et identiques aux comportements actuels).

Enfin, la répartition des internes de première année ayant repassé les ECN à l'issue de leur première année d'internat et ayant été réaffectés peut être modifiée.

## **C – Module 2 – L'entrée dans la vie active**

Le module précédent produit pour une année donnée de la période de projection un fichier de données individuelles, composé d'un enregistrement pour chaque diplômé au cours de l'année courante, contenant comme informations la spécialité et la région (de diplôme) du jeune diplômé.

Le module 2 simule quant à lui l'entrée dans la vie active des jeunes diplômés. Il commence par sélectionner les diplômés de l'année courante qui exerceront la médecine et affecter un délai à l'entrée dans la vie active à chacun d'eux. Il rassemble ensuite les médecins diplômés au cours de l'année courante de la période de projection et ceux diplômés au cours des années précédentes entrant dans la vie active au cours de l'année courante, puis attribue à chacun d'eux successivement un sexe et un âge, une région d'exercice, un mode d'exercice et une zone d'exercice. Enfin, il les fait entrer dans la vie active, c'est-à-dire qu'il les intègre aux effectifs de médecins en activité.

## ***1. Sous-module 2.1 : le calcul des effectifs de médecins qui exerceront la médecine et l'attribution d'un délai à l'entrée dans la vie active***

### ***a. Le « taux de fuites » ou le taux de médecins qui n'exerceront jamais la médecine***

Nous considérons que la date d'entrée dans la vie active est la date d'inscription à l'Ordre des médecins, qui marque le début de l'exercice, puisque tout médecin se proposant d'exercer la médecine (hormis les médecins des armées), est tenu de s'inscrire au tableau de l'Ordre (article L 4112-6 du code de la santé publique).

Certains diplômés entrent dans la vie active l'année (civile) même de l'obtention de leur diplôme de docteur en médecine (DES), d'autres une ou plusieurs années après. Enfin, certains n'exercent jamais la médecine, parce qu'ils décèdent après l'obtention de leur diplôme sans s'être inscrits à l'Ordre, parce qu'ils partent (exercer ou non) à l'étranger, parce qu'ils restent inactifs ou exercent un métier ne nécessitant pas une inscription à l'Ordre (par exemple journaliste médical, fonctionnaire de l'Etat ou agent titulaire d'une collectivité locale n'étant pas appelé, dans le cadre de ses fonctions, à exercer la médecine). On parle, pour ces médecins qui n'exercent jamais, de « fuites ».

On a cherché à estimer « le taux de fuites ». Les effectifs de diplômés qui n'entrent jamais dans la vie active ne peuvent être estimés que par différence entre les effectifs de diplômés et les effectifs de diplômés qui entrent dans la vie active. En comptabilisant les médecins diplômés une année  $a$  inscrits à l'Ordre au 31 décembre 2006, on peut obtenir un effectif approché de médecins qui entrent dans la vie active parmi les médecins diplômés au cours de l'année  $a$ , en faisant l'hypothèse que les médecins diplômés au cours de l'année  $a$  sont restés inscrits à l'Ordre (c'est-à-dire qu'ils n'ont pas interrompu leur activité) entre la date de leur inscription à l'Ordre et le 31 décembre 2006. Il faut seulement que ce calcul porte sur les cohortes (années de diplôme) suffisamment anciennes pour que ces diplômés aient « eu le temps » d'entrer dans la vie active. Cependant, on ne dispose pas de données fiables sur les effectifs de diplômés. On ne peut donc estimer correctement le taux de fuites.

La version précédente du modèle retenait un taux de fuites de 3 % pour les hommes et de 4 % pour les femmes. Le Centre de Sociologie et de Démographie Médicale admettait quant à lui que le taux de fuites était de 4 % pour l'exercice de projections qu'il a réalisé en 2004<sup>11</sup>. En l'absence d'informations complémentaires, on a choisi de faire l'hypothèse d'un taux de fuites égal à 3 %, sans distinction entre hommes et femmes (et sans autre distinction). Ce taux de fuites est appliqué aux effectifs de diplômés de chacune des années de la période de projection.

Le taux de fuites retenu est un paramètre du modèle. Sa valeur peut être modifiée.

Les jeunes diplômés de l'année qui n'exerceront jamais la médecine sont sélectionnés parmi les diplômés de l'année (sortie du sous-module 1.3), par un tirage aléatoire (voir A. 1. d.).

### ***b. L'attribution d'un délai à l'entrée dans la vie active***

La répartition des diplômés des années passées entrés dans la vie active (répertoriés dans le dernier fichier de l'Ordre au 31 décembre 2006) par cohorte (année de diplôme) et délai à l'entrée dans la vie active (écart entre l'année de l'entrée dans la vie active et année de l'obtention du diplôme, en années) montre que la grande majorité des jeunes diplômés qui exercent la médecine s'inscrivent à l'Ordre dans les cinq années qui suivent l'année de l'obtention de leur diplôme. Par exemple, parmi les jeunes diplômés de 1996, diplômés depuis 10 ans, seulement 0,8 % se sont installés après 5 années (donc

---

<sup>11</sup> « Projection démographique de la profession médicale en France (2000-2050) », cahier de sociologie et de démographie médicales, janvier-mars 2004, p101-148.



entre 6 et 10 années après l'obtention de leur diplôme). Pour les cohortes plus anciennes, ce pourcentage n'est pas beaucoup plus élevé, bien que ces cohortes aient eu plus de temps pour commencer à exercer. Il atteint par exemple 2,6 % pour les diplômés de 1981, qui ont eu 25 ans pour s'inscrire à l'Ordre.

Pour le calcul des probabilités d'entrer dans la vie active après un délai  $d$ , il est donc possible de faire l'hypothèse que les médecins commencent à exercer la médecine dans les 5 années suivant l'obtention de leur diplôme. Dès lors, les probabilités d'entrer dans la vie active l'année du diplôme, un an après, deux ans après, etc. doivent être estimées sur le champ des médecins diplômés 5 ans au moins avant la date d'observation, c'est-à-dire, dans le cas présent, diplômés en 2001 ou avant. En effet, tous ces médecins ont eu au moins 5 ans pour s'inscrire à l'Ordre (ce qui n'est pas le cas des médecins diplômés après 2001), et on fait l'hypothèse que ceux d'entre eux qui ne sont pas encore inscrits à l'Ordre n'exerceront jamais la médecine.

Réduire le délai maximal entre l'obtention du diplôme et l'entrée dans la vie active permet d'estimer les probabilités d'entrer dans la vie active après un délai  $d$  à partir du comportement observé de cohortes plus récentes. Si le délai maximal était fixé à 10 ans, il aurait fallu estimer les probabilités à partir du comportement des médecins diplômés en 1996 ou avant, qui est peut-être différent des comportements plus récents.

Ainsi, les probabilités sont calculées par rapport à l'effectif total de médecins diplômés au cours d'une année donnée et entrés dans la vie active dans les 5 années suivantes.

Ces probabilités sont estimées à partir du fichier de l'Ordre au 31 décembre 2006, fichier dans lequel ne seront pris en compte que les médecins ayant obtenu un diplôme d'Etat de docteur en médecine en France, puisque les probabilités estimées seront appliqués à des diplômés en France.

Les trois quarts des médecins exerçant la médecine entrent dans la vie active l'année même de l'obtention de leur diplôme. C'est ce que fait apparaître la répartition des diplômés en France au cours de l'année 2001 (ou d'une année antérieure) entrés dans la vie active dans les 5 années suivant l'obtention de leur diplôme selon leur délai à l'entrée dans la vie active. Au contraire, les médecins s'inscrivant à l'Ordre plus de 2 ans après l'obtention de leur diplôme sont très peu nombreux. Ainsi parmi les médecins diplômés en France en 2001 et inscrits à l'Ordre au cours des 5 années suivantes, seulement 0,6 % se sont inscrits 3 ans après l'obtention de leur diplôme, 0,3 % se sont inscrits 4 ans après et 0,2 % 5 ans après. Ces proportions étant faibles et devant être appliquées à des effectifs assez petits<sup>12</sup>, il serait difficile, dans la simulation, de sélectionner de manière satisfaisante les médecins à installer 3, 4 ou 5 ans après l'obtention de leur diplôme. On a donc choisi de regrouper les probabilités d'entrer dans la vie active 2, 3, 4 ou 5 ans après l'obtention du diplôme pour constituer une probabilité d'entrer dans la vie active 2 ans ou plus après l'obtention du diplôme. Au cours de la simulation, le modèle fait entrer dans la vie active 2 ans après l'obtention du diplôme les médecins qui auraient dû commencer à exercer la médecine 2, 3, 4 ou 5 ans après l'obtention de leur diplôme. Ce faisant, il fait donc commencer à exercer plus tôt qu'il l'aurait théoriquement fallu 1 à 2 % des médecins d'une cohorte qui exerceront la médecine. Cette hypothèse n'est donc pas une hypothèse forte.

Les répartitions des médecins exerçant la médecine selon le délai d'entrée dans la vie active sont calculées pour plusieurs années de diplôme, le nombre d'années prises en compte étant paramétrable. Il n'est pas fixé de maximum pour le nombre d'années de diplôme retenues, mais il est préférable de ne pas retenir des cohortes trop anciennes, dont le comportement pourrait être très différent de celui des générations présentes et futures. Pour l'exercice de projections de 2008, on a retenu 5 cohortes (1997 à 2001). Les probabilités estimées correspondent à la moyenne des répartitions calculées sur les

---

<sup>12</sup> L'effectif total de diplômés est de l'ordre de quelques milliers, mais les probabilités peuvent être appliquées à des effectifs bien plus faibles si elles sont calculées par exemple pour chaque spécialité.

différentes cohortes. Elles sont maintenues constantes pendant toute la durée de la période de projection, sauf si le scénario simulé ne retient pas l'hypothèse de comportements constants en ce qui concerne le choix du délai à l'entrée dans la vie active.

Chaque « événement » du module 2 « entrée dans la vie active » (choix d'un délai à l'entrée dans la vie active, choix d'un sexe et d'un âge, choix d'une région d'exercice, choix d'un mode d'exercice et choix d'une zone d'exercice) et des modules 3 et 4 (« changements au cours de la vie active » et « sortie de la vie active ») a fait l'objet d'une étude qui a cherché à déterminer les critères (parmi ceux pris en compte dans le cadre des projections) qui le « déterminaient » (le sens de la causalité n'est pas établi) le plus (voir encadré 3).

Ainsi, on a cherché à savoir si le délai à l'entrée dans la vie active était nettement corrélé ou non à la spécialité et à la région de diplôme. Seuls ces deux facteurs potentiels ont été étudiés puisqu'au moment où le modèle affecte le délai à l'installation aux futurs médecins, ce sont les deux seules informations disponibles sur les futurs entrants.

Le délai à l'entrée dans la vie active n'est pas indépendant de la spécialité. Il n'est pas non plus indépendant de la région de diplôme. Cependant, d'une part, la corrélation entre le délai à l'entrée dans la vie active et la spécialité (à région de diplôme donnée) semble plus forte que la corrélation entre le délai et la région de diplôme (à spécialité donnée) et d'autre part, il paraît souhaitable, pour les résultats produits, que le lien entre le délai à l'entrée dans la vie active et la spécialité soit pris en compte par le modèle, puisque sa prise en compte influe (plutôt faiblement, par le report ou l'avancement de l'entrée des jeunes diplômés dans la vie active) sur les effectifs de médecins actifs par spécialité au cours de chaque année de la période de projection. Il n'est pas possible de tenir compte à la fois de la spécialité et de la région de diplôme, les strates ainsi obtenues étant trop petites (voir encadré 3). On a donc choisi de calculer les probabilités d'entrer dans la vie active après un délai  $d$  pour chaque spécialité, et de ne pas retenir le critère région de diplôme pour l'estimation des probabilités. Toutefois, les critères retenus constituent un paramètre du modèle.

Le délai à l'entrée dans la vie active est attribué à chaque médecin diplômé au cours de l'année qui exercera la médecine, par un tirage aléatoire (voir A. 1. d.), en fonction des probabilités estimées pour chaque strate constituée par les critères retenus. Si la probabilité correspondant à une strate n'a pas été calculée, cette strate n'ayant pas été observée dans les données ayant servi à estimer les probabilités d'entrer dans la vie active après tel ou tel délai, alors le modèle attribue un délai à l'entrée dans la vie active égal à zéro (modalité la plus fréquente) aux individus de la strate.

### Encadré 3

#### Le choix des critères retenus pour le calcul des probabilités de vivre un événement (modules 2, 3 et 4)

Pour le module 1, simulant le déroulement des études médicales, la question s'est posée de savoir s'il fallait prendre compte la région et la discipline ECN dans le calcul des taux de redoublements, d'abandons, etc., c'est-à-dire s'il était préférable de calculer des taux nationaux ou régionaux, des taux globaux ou par discipline. Les effectifs sur lesquels étaient calculés ces taux, et la variabilité des taux d'une région à l'autre et d'une discipline à l'autre, notamment, ont permis de trancher. En revanche, le niveau d'études était un critère à prendre impérativement en compte, les taux devant être calculés pour chaque niveau d'études.

Pour les autres modules, la question du choix des critères à retenir pour la modélisation de chacun des événements est de même nature, mais plus complexe. Prenons l'exemple de la modélisation du changement de mode d'exercice. Le modèle sélectionne d'abord les médecins qui changent de mode d'exercice entre l'année  $a$  et l'année  $a+1$  (étape 1), puis il affecte aux médecins qui changent de mode d'exercice un nouveau mode d'exercice (étape 2). Le modèle utilise donc pour la première étape des probabilités de changer de mode d'exercice, et pour la seconde des probabilités de choisir comme nouveau mode d'exercice le mode  $m$ . Intéressons-nous aux probabilités de changer de mode d'exercice utilisées dans la première étape.

Une étape préalable à la programmation du sous-module correspondant (3.2) a consisté à déterminer les caractéristiques des médecins les plus liées au fait de changer ou non de mode d'exercice et à réfléchir à l'intérêt, par rapport aux résultats attendus des projections, de chercher à conserver la relation entre chacune des caractéristiques et la répartition des médecins selon le mode d'exercice (voir D. 2.). Il s'agissait en fait de repérer les critères à prendre en compte pour le calcul des probabilités et pour la sélection, au cours de la simulation, des médecins changeant de mode d'exercice.

Il semble que le mode d'exercice « d'origine » (en a), l'âge et la spécialité soient les critères à prendre en compte en priorité. En effet :

- Par rapport aux médecins salariés hospitaliers, les médecins libéraux ont une probabilité de changer de mode d'exercice plus élevée, contrairement aux médecins ayant un exercice mixte, tandis que les médecins salariés non hospitaliers ne se distinguent pas des médecins salariés hospitaliers. Les modèles logistiques ayant servi à modéliser la probabilité de changer de mode d'exercice donnent des paramètres moyennement élevés<sup>13</sup> pour la variable mode d'exercice d'origine, mais on tient à retenir ce critère, afin de garantir un tirage dans des proportions « correctes » des médecins changeant de mode d'exercice au sein de chaque groupe de médecins ayant un mode d'exercice donné, la répartition des médecins par mode d'exercice au cours de chaque année de la période de projection en dépendant très étroitement.

- La probabilité de changer de mode d'exercice décroît avec l'âge. L'âge semble très corrélé au fait de changer de mode d'exercice, les paramètres donnés par les mêmes modèles logistiques étant assez élevés<sup>14</sup>. De plus, il est important de garantir un tirage dans des proportions « correctes » des médecins changeant de mode d'exercice au sein d'un groupe de générations donné, la répartition des médecins par mode d'exercice en dépendant indirectement, via les cessations d'activité (notamment les cessations définitives d'activité), nettement plus fréquentes aux âges avancés. Ainsi, supposons que l'on tire trop de médecins âgés et pas assez de médecins jeunes parmi les libéraux pour les faire changer de mode d'exercice. Ces médecins sont ensuite répartis dans les trois autres modes d'exercice. Or, parmi ceux-ci, la proportion de ceux qui cesseront leur activité dans les années suivantes sera plus élevée que celle attendue, du fait de leur moyenne d'âge également plus élevée que celle attendue. Les effectifs de médecins pour chacun des modes d'exercice se trouveraient donc indirectement affectés par le défaut du tirage.

- Toutes choses égales par ailleurs, les spécialistes se distinguent peu des généralistes en ce qui concerne la probabilité de changer de mode d'exercice. En revanche, les paramètres des modèles logistiques relatifs à la spécialité, bien que modérément élevés, font apparaître que la probabilité de changer de mode d'exercice est plus élevée pour les endocrinologues, les chirurgiens, les ORL, les biologistes et les psychiatres et moins élevée pour les médecins du travail et internistes que pour les gastroentérologues, qui servent de référence. Or, nous jugeons important que les résultats des projections permettent de donner un aperçu assez précis des effectifs de chaque spécialité par mode d'exercice. La spécialité est donc un critère à retenir, bien que seulement quelques spécialités se distinguent un peu de la spécialité de référence.

- La région d'exercice apparaît peu discriminante pour le changement de mode d'exercice. Les paramètres associés aux modalités de cette variable sont modérément élevés, voire faibles. Les effets que fait apparaître la régression pour certaines régions ne sont pas stables dans le temps : ils n'apparaissent pas systématiquement pour chaque année et sont même dans certains cas inversés d'une année observée à l'autre pour une même région. Enfin, il paraît souhaitable, mais pas primordial, que les résultats des projections permettent d'évaluer précisément les effectifs régionaux de médecins futurs par mode d'exercice.

- Le fait de changer de mode d'exercice est faiblement corrélé au sexe du médecin et à sa zone d'exercice. Il ne semble donc pas indispensable de retenir ces critères.

- Enfin, il n'importe pas de préserver la relation entre le fait de changer de mode d'exercice et le fait de changer de région au cours de l'année, bien que celle-ci soit forte. En effet, les projections visent à évaluer les effectifs de médecins par région d'exercice, par mode d'exercice, par âge, etc., ces critères pouvant éventuellement être croisés, mais pas à étudier les comportements des médecins en matière de mobilité. Il est donc indifférent, pour les résultats attendus des projections, que le médecin qui change de mode d'exercice soit ou non le même que celui qui change de région, même si la probabilité de changer de mode d'exercice et la probabilité de changer de région ne sont pas indépendantes dans la réalité.

Toutefois, il aurait pu s'avérer nécessaire de prendre en compte le lien entre le changement de région et le changement de mode d'exercice si l'on avait souhaité retenir la région d'exercice comme critère pour le calcul des probabilités. En effet, le modèle simule les changements de région avant les changements de mode d'exercice. Au moment où il doit simuler les changements de mode d'exercice, la nouvelle région d'exercice des médecins ayant changé de région est donc connue. Or, les probabilités de changer de mode d'exercice seraient en principe calculées par rapport à l'ancienne région d'exercice, puisqu'on estime les probabilités en observant les changements intervenus entre l'année a et l'année a+1. Pour la simulation, si l'on retenait le critère région, il faudrait donc veiller à ce que le modèle se réfère bien à l'ancienne région d'exercice. Mais ce faisant, on ne contrôlerait pas la répartition des médecins changeant de mode d'exercice en fonction de la nouvelle région d'exercice. A défaut de tenir compte, pour les médecins ayant changé de région, de l'ancienne région et de la nouvelle région, on aurait pu retenir comme critère, en plus de l'ancienne région, le fait d'avoir ou non changé de région et faire l'hypothèse que la répartition selon la nouvelle région d'exercice des médecins que le modèle sélectionnerait pour les faire changer de région serait semblable à celle observée sur les données ayant servi au calcul des probabilités.

L'étape préalable à la programmation du sous-module a donc permis d'effectuer une première sélection parmi les critères qui pourraient être à retenir pour le calcul des probabilités de changer de mode d'exercice. Dans cette étape, les critères sélectionnés doivent être ceux dont les effets sont vraiment nets et stables d'une année à l'autre, ou ceux dont l'intérêt par rapport aux résultats attendus de l'étude est important. En effet, nous allons voir que la prise en compte d'un critère n'est pas neutre sur les probabilités estimées et sur le résultat des tirages qui utilisent ces probabilités.

<sup>13</sup> Pour les changements de mode d'exercice intervenus entre 2005 et 2006, dans le modèle comprenant comme variables explicatives le mode d'exercice, la spécialité, l'âge et le sexe, avec le mode « salarié hospitalier » comme modalité de référence, le paramètre est de -0,69 pour la modalité « libéral », de 0,36 pour la modalité « salarié non hospitalier » et non significativement différent de zéro pour la modalité « mixte ». Les paramètres sont un peu plus faibles pour les changements entre 2004 et 2005 et entre 2003 et 2004.

<sup>14</sup> Dans le modèle décrit dans la note précédente, pour les changements intervenus entre 2005 et 2006, avec la tranche d'âge « 25-35 ans » comme modalité de référence, les paramètres pour la variable âge sont compris entre -0,58 et -1,48.

Au vu des éléments présentés au paragraphe précédent, on est tenté de calculer des probabilités de changer de mode d'exercice pour chaque strate composée par le croisement des variables mode d'exercice, âge et spécialité entre elles, dans la mesure où ces trois variables étant corrélées au fait de changer de mode d'exercice, les probabilités ainsi obtenues seraient probablement différentes d'une strate à l'autre. Cependant, plusieurs éléments sont encore à peser avant de retenir définitivement ces trois critères.

Tout d'abord, se pose la question des effectifs sur lesquels sont calculées les probabilités correspondant à chaque strate. Dans le cas présent, ces effectifs sont plutôt conséquents dans la plupart des strates, mais assez faibles, voire très faibles pour les strates correspondant au mode d'exercice « salarié non hospitalier ». Les faibles effectifs font craindre premièrement une forte volatilité d'une année à l'autre des taux qu'ils servent à calculer, que le calcul de taux moyens ne pourrait corriger correctement. Deuxièmement, si les strates regroupent peu de médecins dans les données qui servent à estimer les probabilités, elles regrouperont sans doute également peu de médecins dans les données à partir desquelles les tirages seront réalisés au cours de la simulation, ce qui ne favorisera pas la bonne réalisation de ces tirages. Toutefois, même si les taux ne sont pas très stables pour certaines strates, les faibles effectifs ne posent pas toujours de gros problème et sont parfois « acceptables », ce qui semble *a priori* le cas ici : ils ne font pas apparaître à première vue l'estimation des probabilités comme trop risquée, les faibles effectifs ne concernant que certaines strates et les taux correspondant à ces strates étant « raisonnablement » volatiles.

Se pose ensuite la question du gain que l'on retire de l'ajout d'un critère. *A priori*, on peut penser que plus on prend en compte de critères, plus le tirage sera fidèle à la réalité, sauf si les effectifs observés dans chaque strate deviennent trop petits. Tenir compte de l'âge par exemple pour la sélection des médecins changeant de mode d'exercice permet de tenir compte de la corrélation entre l'âge et la propension à changer de mode d'exercice, et de « garantir » que la répartition des médecins qui changeront de mode d'exercice en fonction de leur âge correspondra bien à celle observée en réalité. Cela permet également d'améliorer la précision du modèle dans la mesure où, comme on l'a vu, les effectifs correspondant à chaque mode d'exercice sont indirectement liés à la prise en compte de ce critère pour la simulation, via les cessations d'activité notamment. On peut donc être tenté d'ajouter les critères un par un dans leur ordre d'intérêt et de s'arrêter dès que les effectifs deviennent manifestement trop petits, en particulier lorsque les critères sont très hiérarchisés (c'est-à-dire lorsque l'intérêt de les prendre en compte est très inégal).

Mais en réalité, on ne gagne pas systématiquement à ajouter un critère supplémentaire, pour deux raisons :

- Si la corrélation entre la probabilité de réalisation de l'événement que l'on cherche à simuler et le critère à ajouter n'est pas particulièrement forte, il est fort probable que les résultats du tirage seront décevants. Par exemple, pour le critère de l'âge pour la modélisation du fait de changer ou non de mode d'exercice, le modèle pourrait (dans le cas d'une corrélation faible ou moyenne) n'améliorer que très faiblement la répartition par âge des médecins sélectionnés pour changer de mode d'exercice et pourrait ne produire des résultats meilleurs que pour certaines strates, voire produire pour certaines strates des résultats moins bons (éventuellement nettement moins bons) que ceux obtenus sans ce critère supplémentaire.

Par exemple, lorsqu'on n'utilise comme critère que le mode d'exercice, on sélectionne, avec une racine donnée pour initialiser la procédure ranuni de SAS sur un fichier test donné, un effectif de médecins hospitaliers de 25 à 35 ans devant changer de mode d'exercice de 385, alors que l'effectif-cible, déduit de la probabilité estimée et de l'effectif de la strate, est de 752. Lorsqu'on ajoute l'âge comme critère, en plus du mode d'exercice, l'effectif obtenu pour cette strate est de 751. Le résultat du tirage est donc nettement meilleur pour cette strate avec ce critère supplémentaire. En revanche, on constate le contraire pour les hospitaliers de 41 à 45 ans, pour lesquels l'effectif-cible est de 281, l'effectif obtenu avec le mode d'exercice comme seul critère de 286 et l'effectif obtenu avec le mode d'exercice et l'âge comme critères 252. Dans le cas où la corrélation est particulièrement forte, les tests réalisés montrent que l'ajout d'un critère supplémentaire améliore en général le résultat du tirage au regard de ce critère, mais malgré tout, du fait de l'aléa inhérent au tirage et éventuellement de la petite taille de certaines strates, il est extrêmement rare que les résultats du tirage s'avèrent meilleurs pour toutes les strates. Il faut donc juger de l'intérêt de l'ajout d'un critère supplémentaire en évaluant les résultats obtenus pour l'ensemble des strates pour déterminer si l'ajout du critère supplémentaire est « globalement » utile, en comparant par exemple les écarts quadratiques aux probabilités-cibles.

- L'ajout d'un critère supplémentaire peut détériorer les résultats obtenus sans ce critère en termes de répartition des individus tirés selon les critères déjà pris en compte.

Par exemple, d'après l'effectif de médecins hospitaliers et la probabilité estimée correspondante, l'effectif-cible de médecins hospitaliers à sélectionner pour changer de mode d'exercice est de 1 852. Le tirage effectué avec le mode d'exercice comme seul critère donne un effectif de médecins hospitaliers sélectionnés de 1862, tandis que celui effectué avec le mode d'exercice et l'âge comme critères aboutit à un effectif de 1758, moins proche de l'effectif-cible.

Pour évaluer la qualité des sélections obtenues en prenant ou non en compte chacun des critères que l'on envisage de retenir, il est possible également de réaliser la simulation sur les données qui ont servi à estimer les probabilités utilisées en entrée du modèle (sur les données d'une année observée), et de comparer les résultats obtenus par la simulation à la réalité.

Finalement, pour sélectionner les critères à prendre en compte, il faut s'interroger sur :

- la force du lien entre ce critère et la probabilité que l'événement à modéliser se réalise et sa stabilité dans le temps
- l'intérêt pour les résultats attendus des projections de préserver ce lien dans la simulation
- la taille des strates obtenues selon que l'on tient compte ou non de ce critère et ses conséquences
- le gain en ce qui concerne la répartition des individus sélectionnés selon chaque critère

et effectuer un arbitrage qui tiennent compte de l'ensemble de ces éléments.

Les résultats obtenus sont dépendants de l'aléa généré par le tirage, ou plus précisément des valeurs de la simulation de la réalisation de la loi uniforme qui sert au tirage. Ils varient donc selon que l'on utilise telle ou telle racine pour initialiser la procédure ranuni de SAS. Or, le modèle utilise plusieurs racines pour une même simulation, et en réalité, toutes les racines, dont le nombre est infini (ou

presque), peuvent être utilisées. Il faut donc, au vu des résultats obtenus en réalisant un ou quelques tests (sur la base d'une ou quelques racines), se faire une idée des résultats qui pourraient être produits quelle que soit la racine utilisée.

Les résultats, enfin, dépendent de la façon dont est programmé le tirage : par exemple, si l'on note, pour deux événements simultanés A et B, probA la probabilité que l'individu *i* vive l'événement A, probB la probabilité qu'il vive l'événement B et *alea* la valeur de la loi uniforme générée pour chaque individu, les résultats obtenus différeront selon que l'on sélectionne les individus qui vivront les événements A et B comme ceci :

```
if alea <= probA then evenement = 'A' ;  
else if alea > probA and alea <= (probA+probB) then evenement = 'B' ;  
ou comme suit :  
if alea <= probB then evenement = 'B' ;  
else if alea > probB and alea <= (probA+probB) then evenement = 'A' ;  
bien que ces deux méthodes soient correctes.
```

### ***c. Le cas particulier des médecins diplômés au cours de l'une des cinq années précédant le début de la période de projection***

Les médecins diplômés à partir de l'année de début de la période de projection qui exerceront la médecine seront tous intégrés à la base des médecins en activité par le modèle, plus ou moins rapidement, en fonction du délai à l'entrée dans la vie active que le modèle lui-même leur aura attribué (sauf une partie des médecins diplômés dans les deux dernières années de la période de projection).

Au contraire, certains des médecins diplômés au cours de l'une des cinq années précédant le début de la période de projection ont déjà pu s'inscrire à l'Ordre. Dans ce cas, ils sont présents dans la base des médecins en activité au 31/12/2006 (fichier de l'Ordre) qui sert de base à la simulation, et ne doivent pas être à nouveau intégrés à la base des médecins en activité par le modèle, pour ne pas créer de doubles-comptes. Au contraire les médecins diplômés au cours de l'une des cinq années précédant le début de la période de projection non encore entrés dans la vie active au début de la période de projection mais qui exerceront la médecine doivent être progressivement intégrés par le modèle à la base des médecins en activité.

Prenons le cas des médecins diplômés en 2004. Parmi ces médecins, certains ont pu entrer dans la vie active en 2004, en 2005 ou en 2006, et sont donc présents dans la base initiale de médecins en activité utilisée par le modèle (la base du CNOM au 31/12/2006 pour l'exercice de projections de 2008). Ceux qui exerceront la médecine et qui ne se sont pas encore inscrits à l'Ordre au 31/12/2006 peuvent encore le faire en 2007, en 2008 et en 2009 (année de l'obtention de leur diplôme +5 ans). Les premiers sont déjà présents dans la base des médecins en activité initiale, tandis que les seconds doivent être intégrés par le modèle à la base des médecins en activité.

Si l'on connaissait précisément le nombre de diplômés en 2004, on pourrait en déduire le nombre de diplômés en 2004 qui exerceront la médecine en ôtant de cet effectif de diplômés le nombre de « fuites ». À défaut de données fiables sur le nombre de diplômés en 2004, on estime le nombre de diplômés en 2004 qui exerceront la médecine à partir du nombre de ces diplômés déjà inscrits à l'Ordre au 31/12/2006.

Pour ce faire, on calcule dans un premier temps, à partir du fichier de l'Ordre au 31/12/2006, les proportions de jeunes diplômés entrés dans la vie active l'année de l'obtention du diplôme, 1 an, 2 ans, 3 ans, 4 ans ou 5 ans après. Ces proportions sont estimées de la même façon que les probabilités d'entrer dans la vie active l'année du diplôme, 1 an ou 2 ans ou plus après, probabilités qui seront appliquées aux médecins diplômés à partir du début de la période de projection (voir C. 1. b.). Soient moydelai0 la proportion de diplômés entrés dans la vie active l'année de l'obtention du diplôme, moydelai1 la proportion de diplômés entrés dans la vie active 1 an après l'obtention du diplôme,... et moydelai5 la proportion de diplômés entrés dans la vie active 5 ans après l'obtention du diplôme.

Soient par ailleurs  $n_0$  le nombre de diplômés de 2004 entrés dans la vie active l'année de l'obtention du diplôme (donc en 2004) pour une strate donnée,... et  $n_5$  le nombre de diplômés de 2004 entrés dans la vie active 5 ans après l'obtention du diplôme, c'est-à-dire en 2009, pour cette même strate.

Le modèle calcule alors  $nb_{dip}$ , le nombre de diplômés de 2004 qui exerceront la médecine comme suit (par une règle de trois), pour cette strate :

$$nb_{dip} = (n_0+n_1+n_2) / (\text{moydelai}_0+\text{moydelai}_1+\text{moydelai}_2)*100^{15}$$

Il calcule ensuite le nombre de diplômés de 2004 à faire entrer dans la vie active en 2007 ( $n_3$ ), en 2008 ( $n_4$ ) et en 2009 ( $n_5$ ) pour la strate en appliquant à  $nb_{dip}$  les proportions calculées :

$$n_3 = nb_{dip}*(\text{moydelai}_3/100)$$

$$n_4 = nb_{dip}*(\text{moydelai}_4/100)$$

$$n_5 = nb_{dip}*(\text{moydelai}_5/100)$$

Ces calculs sont effectués selon les critères retenus (ces critères sont les mêmes que ceux retenus pour l'estimation des délais à l'entrée dans la vie active pour les médecins diplômés à partir de la première année de la période de projection – pour les projections de 2008, la spécialité a été retenue), afin de tenir compte de la propension à entrer dans la vie active plus ou moins rapidement selon les critères retenus. Ils sont effectués également par région de diplôme, afin que la répartition des médecins entrant dans la vie active par région de diplôme ne soit pas trop déformée d'une part, et pour attribuer une région de diplôme à ces nouveaux médecins d'autre part. (En pratique, les proportions appliquées sont calculées selon les critères retenus et appliquées à chaque région.)

Le modèle procède aux mêmes calculs pour les diplômés de 2006, de 2005, de 2003 et de 2002 (année de début de la période de projection – 5 ans).

Enfin, pour chacun des diplômés de ces 5 dernières années, le modèle crée un enregistrement contenant l'année de diplôme, la spécialité, la région de diplôme et l'année d'entrée dans la vie active (qui est attribuée cette fois sans tirage aléatoire, en fonction des effectifs  $n_0$ ,  $n_1$ ,  $n_2$ ,  $n_3$ ,  $n_4$  et  $n_5$  calculés). Certains de ces médecins entreront dans la vie active au cours de la simulation.

#### ***d. Création de la base des médecins à faire entrer dans la vie active au cours de l'année***

Pour une année donnée de la période de projection, après avoir sélectionné ceux des médecins diplômés au cours de l'année qui entreront dans la vie active (c'est-à-dire ôté « les fuites »), à l'aide d'un tirage aléatoire en fonction du taux de fuites, et attribué à ceux des médecins diplômés au cours de l'année qui exerceront la médecine un délai à l'entrée dans la vie active, le modèle dispose d'une base comprenant un enregistrement par diplômé au cours de l'année qui exercera la médecine, chaque enregistrement contenant comme informations l'année de diplôme, la région de diplôme, la spécialité et l'année d'entrée dans la vie active.

Il place ces enregistrements dans une base dite « des médecins attendant d'entrer dans la vie active », dans laquelle les enregistrements déjà présents correspondent aux médecins diplômés au cours de l'une des deux années précédentes (ajoutés dans cette base au cours de l'un des deux cycles précédents) ou au cours de l'une des 5 années précédant le début de la période de projection et devant encore entrer dans la vie active (ajoutés au cours du premier cycle du modèle - correspondant à l'année de début de la période de projection).

---

<sup>15</sup> On fait l'hypothèse que si  $n_0+n_1+n_2$  est nul, alors  $nb_{dip}$  est nul, ce qui est une hypothèse faible car la majorité des jeunes diplômés entrent dans la vie active l'année de l'obtention de leur diplôme.

En outre, si  $\text{moydelai}_0$  (et par conséquent  $\text{moydelai}_1$  et  $\text{moydelai}_2$ ) n'a pu être estimée, la strate n'ayant pas été observée dans les données ayant servi à l'estimation des probabilités, on suppose que  $nb_{dip} = n_0+n_1+n_2$ .

Le modèle sélectionne alors dans la base des médecins attendant d'entrer dans la vie active ceux des médecins devant entrer dans la vie active au cours de l'année.

Le résultat de cette étape est donc une base de données individuelles, la base des médecins à faire entrer dans la vie active au cours de l'année, contenant comme informations, pour chaque enregistrement, l'année de diplôme, la région de diplôme, la spécialité et l'année d'entrée dans la vie active. Est créée également la variable *genspe*, qui distingue les généralistes des spécialistes. Cette table contient enfin la variable *delaiins*, le délai à l'entrée dans la vie active. Les enregistrements correspondent aux diplômés de l'année qui entrent dans la vie active au cours de l'année de l'obtention de leur diplôme, les diplômés de l'année précédente qui entrent dans la vie active 1 an après l'obtention de leur diplôme, les diplômés d'il y a 2 ans qui entrent dans la vie active 2 ans ou plus après l'obtention de leur diplôme et les diplômés de l'une des 5 années précédant le début de la période de projection qui entrent dans la vie active au cours de l'année courante.

Le modèle supprime de la base des médecins attendant d'entrer dans la vie active ceux qu'il fera entrer dans la vie active au cours de l'année, mais ceux-ci ne sont pas immédiatement intégrés à la base des médecins en activité : au préalable, un sexe et un âge, une région, un mode et une zone d'exercice leur sont attribués.

## ***2. Sous-module 2.2 : attribution d'un sexe et d'un âge aux médecins entrant dans la vie active au cours de l'année***

À chaque médecin entrant dans la vie active au cours de l'année sont d'abord attribués un sexe et un âge, simultanément, afin de bien tenir compte de la corrélation existant entre ces deux variables<sup>16</sup>. Cette information est ajoutée à la base des médecins à faire entrer dans la vie active au cours de l'année, créée par le sous-module 2.1.

Pour des raisons de commodité, on estime les **probabilités d'avoir un sexe et un âge au moment de l'obtention du diplôme**, et non au moment de l'entrée dans la vie active.

Ces probabilités sont calculées à partir des fichiers de l'Ordre, à l'aide des variables sexe et âge à l'obtention du diplôme.

Le champ retenu pour le calcul des probabilités est celui auquel elles seront ensuite appliquées : les médecins ayant obtenu un diplôme d'Etat de docteur en médecine en France et entrés dans la vie active dans les 5 années ayant suivi l'obtention de leur diplôme.

Le modèle attribuera un âge au diplôme compris entre 26 et 40 ans aux jeunes médecins qu'il fera entrer dans la vie active. L'âge des médecins qui auraient déclaré lors de leur inscription à l'Ordre avoir au moment de leur diplôme moins de 26 ans ou plus de 40 ans (peu nombreux) est donc redressé (à 26 ans ou 40 ans) pour le calcul des probabilités.

Le nombre d'années de diplôme utilisées pour le calcul des probabilités est paramétrable. Comme pour le délai à l'entrée dans la vie active, il est préférable que les cohortes choisies ne soient pas trop anciennes, même si aucun nombre maximal d'années n'est fixé. Pour chaque année de diplôme est calculée la répartition par sexe et âge des jeunes diplômés entrés dans la vie active selon les critères retenus (voir ci-après). Les probabilités estimées correspondent à la moyenne de ces répartitions. Pour

---

<sup>16</sup> Il n'est pas indispensable d'attribuer simultanément le sexe et l'âge des médecins à faire entrer dans la vie active, mais ce choix a été fait pour préserver la corrélation entre ces deux variables pour les médecins entrant dans la vie active. La non prise en compte de cette corrélation pourrait conduire à une mauvaise estimation de la pyramide des âges des médecins en activité au cours de chaque année de la période de projection.

les projections de 2008, 3 années de diplôme (2004 à 2006) ont été utilisées pour le calcul des probabilités d'avoir un sexe et un âge donnés.

Les probabilités sont calculées selon les critères que l'on choisit parmi ceux possibles : la région de diplôme, le fait d'être généraliste ou spécialiste, la spécialité, ou le délai à l'installation (0, 1 ou 2 ans). La région de diplôme et le délai à l'installation n'étant pas des variables d'intérêt pour les résultats, elles n'ont pas été retenues comme critères pour les projections de 2008. L'effet de la spécialité étant net, ce critère a été retenu. Sa prise en compte a un effet sur les résultats en termes d'effectifs par spécialité, l'âge étant corrélé à la probabilité de sortir temporairement ou définitivement de la vie active. Cet effet est toutefois limité puisque d'une part les médecins qui entreront dans la vie active au cours de la période de projection ont une faible probabilité de sortir définitivement de la vie active avant 2030 et d'autre part, les sorties temporaires ne sont pas massives.

Les probabilités appliquées sont égales à celles observées sur la période récente et maintenues constantes pendant toute la durée de la période de projection, sauf si le scénario simulé ne retient pas l'hypothèse de comportements constants en la matière.

Le sexe et l'âge au diplôme sont attribués par le modèle par un tirage aléatoire (voir A. 1. d.). Si la probabilité n'a pas été calculée pour une strate donnée, la strate n'ayant pas été observée dans les données ayant servi au calcul des probabilités, le modèle attribue aux jeunes médecins de cette strate le sexe masculin et leur donne 30 ans comme âge au moment de l'obtention du diplôme.

Le modèle **calcule** enfin **l'âge à l'entrée dans la vie active** de chaque médecin à faire entrer dans la vie active au cours de l'année, **en ajoutant à l'âge au diplôme** qu'il vient de lui attribuer **le délai à l'entrée dans la vie active** qu'il a lui a attribué au cours de l'étape précédente.

L'information est ensuite ajoutée dans la base des médecins à faire entrer dans la vie active au cours de l'année.

### ***3. Sous-module 2.3 : attribution d'une région d'exercice aux médecins entrant dans la vie active au cours de l'année***

Le modèle attribue ensuite à chacun des médecins entrant dans la vie active au cours de l'année courante une région d'exercice, et ajoute cette information à la base contenant les données relatives à ces médecins, créée par le sous-module 2.1 et augmentée par le sous-module 2.2.

Les probabilités pour un jeune médecin de choisir chacune des régions comme région d'exercice au moment de l'entrée dans la vie active sont estimées à partir des données des fichiers de l'Ordre. Ces fichiers donnent, pour une année donnée, la région d'exercice des médecins en activité. Si l'on souhaite estimer les probabilités en observant le comportement de plusieurs cohortes, le nombre de cohortes retenues étant noté  $c$ , il faut donc observer dans chacun des  $c$  derniers fichiers de l'Ordre au 31 décembre de l'année  $n$  les médecins entrés dans la vie active au cours de l'année  $n$ . En effet, si l'on observe au 31 décembre de l'année  $n$  les médecins entrés dans la vie active au cours d'une année antérieure à  $n$ , il n'est pas certain que la région d'exercice indiquée dans le fichier de l'Ordre corresponde à la région d'exercice au moment de l'entrée dans la vie active, le médecin ayant pu changer de région d'exercice entre le moment de son entrée dans la vie active et le moment de l'observation. Pour estimer les probabilités, on concatène donc les observations des  $c$  derniers fichiers de l'Ordre au 31 décembre de l'année  $n$  correspondant aux médecins installés au cours de l'année  $n$ .

Le nombre de cohortes retenues ( $c$ ) est paramétrable, mais il s'agira toujours des  $c$  cohortes les plus récentes. Toutefois, en 2008, les fichiers de l'Ordre disponibles étaient ceux au 31 décembre des années 2003 à 2006. Ce nombre ne pouvait donc être supérieur à quatre pour les projections de 2008.



Nous avons ainsi utilisé pour l'estimation des probabilités les observations correspondant aux médecins présents dans la base au 31/12/2006 et entrés dans la vie active au cours de 2006, les observations correspondant aux médecins présents dans la base au 31/12/2005 et entrés dans la vie active au cours de 2005, et enfin les observations correspondant aux médecins présents dans la base au 31/12/2004 et entrés dans la vie active au cours de 2004.

Pour chaque cohorte retenue, est calculée la répartition par région d'exercice des jeunes médecins entrés dans la vie active au cours de cette année selon les critères retenus (voir ci-après). Les probabilités estimées correspondent à la moyenne de ces répartitions.

En outre, les probabilités sont calculées sur le champ auquel elles s'appliquent, à savoir les médecins ayant obtenu un diplôme d'Etat de docteur en médecine en France et entrés dans la vie active au cours de l'une des 5 années suivantes.

Les critères pour le calcul des probabilités sont choisis en entrée, où l'on peut sélectionner la région de diplôme, le fait d'être généraliste ou spécialiste, la spécialité, le délai à l'installation (0, 1 ou 2 ans), le sexe et l'âge, informations déjà connues pour les médecins à faire entrer dans la vie active au moment où le modèle doit attribuer une région d'exercice. Concernant l'âge, il s'agit de l'âge à l'entrée dans la vie active, pour lequel on retient seulement 4 modalités : 28 ans ou moins, 29 ans, 30 ans et 31 ans ou plus. Dans le cas où ce critère est retenu, le modèle utilise néanmoins, pour l'estimation des probabilités, à la place de l'âge au moment de la première inscription à l'Ordre, l'âge au diplôme indiqué dans les fichiers de l'Ordre, qu'il redresse s'il est inférieur à 26 ans ou supérieur à 40 ans, et auquel il ajoute le délai à l'entrée dans la vie active (qu'il calcule par différence entre l'âge déclaré à la première inscription à l'Ordre et l'âge déclaré au moment du diplôme), recodé sur 3 modalités : 0, 1 ou 2 ans (pour les mêmes raisons que précédemment – voir sous-module 2.1).

Il apparaît que la corrélation entre la région d'exercice et la région de diplôme est extrêmement forte : les trois quarts des médecins commencent à exercer dans leur région de diplôme. Bien que la région de diplôme ne soit pas une variable d'intérêt pour le résultat des projections, cette information a été retenue comme critère pour le calcul des probabilités, en raison de la force de la corrélation, qui, s'il en est tenu compte pour l'attribution d'une région d'exercice aux nouveaux actifs, doit permettre d'assurer une répartition des entrants dans la vie active par région d'exercice assez proche de celle attendue. Le choix de la région d'exercice à l'entrée dans la vie active diffère par ailleurs nettement selon qu'il s'agit de généralistes ou de spécialistes. Le fait d'être généraliste ou spécialiste a donc été retenu comme critère pour les projections de 2008. Bien que des différences apparaissent également dans les choix des spécialistes en fonction de leur spécialité, cette information n'a pas été retenue comme critère, ce qui aurait conduit à la construction de strates trop petites (encadré 3). Le lien entre la région d'exercice à l'entrée dans la vie active et le sexe s'est avéré négligeable, tandis que le lien entre la région d'exercice à l'entrée dans la vie active et l'âge n'est pas clairement établi. Ces variables n'ont donc pas été retenues comme critères. Enfin, nous n'avons pas tenu compte du délai à l'installation pour le calcul des probabilités, qui ne constitue pas une variable d'intérêt pour les résultats des projections.

Les probabilités sont supposées égales à celles observées au cours des années récentes, et maintenues constantes pendant toute la durée de la période de projection, sauf si le scénario simulé ne retient pas l'hypothèse de comportements constants en ce qui concerne le choix de la région d'exercice à l'entrée dans la vie active.

La région d'exercice à l'entrée dans la vie active est attribuée par le modèle par un tirage aléatoire (voir A. 1. d.). Dans le cas où le modèle doit attribuer une région d'exercice à un entrant faisant partie d'une strate pour laquelle la probabilité n'a pas été calculée (cette strate n'ayant pas été observée dans les données ayant servi à l'estimation des probabilités), il attribue comme région d'exercice à l'entrée dans la vie active la région de diplôme.

L'information est ensuite ajoutée dans la base des médecins à faire entrer dans la vie active au cours de l'année.

#### ***4. Sous-module 2.4 : attribution d'un mode d'exercice aux médecins entrant dans la vie active au cours de l'année***

Le modèle attribue ensuite à chaque médecin entrant dans la vie active au cours de l'année courante un mode d'exercice et ajoute cette information à la base contenant les données relatives à ces médecins, créée par le module 2.1 et augmentée par les modules 2.2 et 2.3.

Les probabilités de choisir un mode d'exercice donné sont estimées à partir des comportements des  $c$  dernières cohortes de médecins entrés dans la vie active, observés dans les  $c$  derniers fichiers de l'Ordre. Comme pour le calcul des probabilités destinées à l'attribution d'une région d'exercice, dans le fichier au 31 décembre de l'année  $n$ , on calcule la répartition par mode d'exercice des seuls médecins entrés dans la vie active au cours de l'année  $n$ , afin de bien observer le mode d'exercice au moment de l'entrée dans la vie active. Cette répartition est calculée pour chacune des  $c$  cohortes retenues, le nombre de cohortes retenues étant paramétrable, mais dans tous les cas, il s'agira des  $c$  cohortes les plus récentes. En 2008, nous disposions des fichiers de l'Ordre de 2003 à 2006. Le nombre de cohortes retenues devait donc être inférieur à quatre. Nous avons retenu les trois dernières. Les probabilités estimées correspondent à la moyenne des répartitions calculées sur les  $c$  cohortes retenues.

Les probabilités de choisir un mode d'exercice donné à l'entrée dans la vie active sont estimées sur un champ identique à celui des médecins auxquelles elles seront appliquées, c'est-à-dire sur le champ des médecins ayant obtenu un diplôme d'Etat de docteur en médecine en France et entrés dans la vie active dans les 5 années ayant suivi l'obtention de leur diplôme.

Les probabilités sont déclinées selon certains critères, choisis parmi ceux qui peuvent être retenus, à savoir la région de diplôme, le délai à l'entrée dans la vie active, la spécialité, le fait d'être généraliste ou spécialiste, le sexe, l'âge à l'entrée dans la vie active (sur 4 modalités, comme pour l'attribution d'une région d'exercice) et la région d'exercice. Les variables région de diplôme et délai à l'entrée dans la vie active, qui ne constituent pas des variables d'intérêt pour les résultats des projections, n'ont pas été retenues comme critères pour les projections de 2008. La spécialité est nettement corrélée au mode d'exercice et l'analyse des résultats des projections pourrait porter sur les effectifs projetés par spécialité et par mode d'exercice. La spécialité a été retenue comme critère. Les corrélations entre le mode d'exercice et l'âge au moment de l'entrée dans la vie active d'une part et le sexe d'autre part sont également marquées. En outre, via les sorties d'activité, l'effet de l'âge sur les effectifs projetés pour chaque mode d'exercice peut s'avérer non négligeable. Enfin, l'effet de la région d'exercice n'est pas neutre, mais seules certaines régions se distinguent des autres et il n'est pas possible de retenir simultanément les critères spécialité et région d'exercice, ce qui donnerait des strates trop petites (encadré 3). Nous avons donc privilégié la spécialité et n'avons pas retenu le critère région d'exercice. Les critères retenus pour le calcul des probabilités de choisir tel ou tel mode d'exercice sont donc finalement la spécialité, le sexe et l'âge à l'entrée dans la vie active. Les strates ainsi obtenues sont de taille assez faible. Toutefois, les résultats obtenus par le tirage aléatoire utilisé pour attribuer un mode d'exercice aux médecins entrant dans la vie active ne sont pas, dans l'ensemble, très éloignés des probabilités-cibles.

Les probabilités sont supposées égales à celles observées au cours des années récentes et maintenues constantes pendant toute la durée de la période de projection, sauf si le scénario simulé ne retient pas l'hypothèse de comportements constants en ce qui concerne le choix du mode d'exercice à l'entrée dans la vie active.

Le modèle affecte un mode d'exercice à chaque médecin à faire entrer dans la vie active en appliquant les probabilités estimées, par un tirage aléatoire (voir A. 1. d.). Lorsqu'il doit attribuer un mode d'exercice à un médecin appartenant à une strate pour laquelle la probabilité n'a pas été calculée (aucun médecin n'ayant appartenu à cette strate dans les fichiers de l'Ordre ayant servi à l'estimation des probabilités), le modèle attribue le mode libéral aux médecins généralistes et le mode salarié hospitalier aux médecins spécialistes.

L'information est ensuite ajoutée dans la base des médecins à faire entrer dans la vie active au cours de l'année.

### ***5. Sous-module 2.5 : attribution d'une zone d'exercice aux médecins entrant dans la vie active au cours de l'année***

Enfin, le sous-module 2.5 du modèle attribue une zone d'exercice aux médecins entrant dans la vie active au cours de l'année courante et ajoute cette information à la base contenant les données relatives à ces médecins.

Les probabilités de choisir une zone d'exercice donnée sont estimées à partir des fichiers de l'Ordre, qui permettent d'observer le comportement des cohortes de médecins entrés récemment dans la vie active. Comme pour l'attribution d'une région et d'un mode d'exercice, on a choisi de retenir un certain nombre de cohortes parmi les cohortes les plus récentes. Pour les projections de 2008, ont été retenues les cohortes de médecins entrés dans la vie active en 2004, 2005 et 2006, pour les mêmes raisons que précédemment. La répartition des médecins de chacune des  $c$  cohortes les plus récentes retenues par zone d'exercice à l'entrée dans la vie active est calculée et les probabilités estimées correspondent à la moyenne des  $c$  répartitions.

Les probabilités sont calculées sur un champ identique au champ des médecins auxquels elles seront appliquées : les médecins ayant obtenu un diplôme de docteur d'État en médecine en France et entrés dans la vie active dans les 5 années ayant suivi l'obtention de leur diplôme.

Les probabilités sont calculées selon les critères choisis parmi la région de diplôme, le délai à l'entrée dans la vie active, la spécialité, le fait d'être généraliste ou spécialiste, le sexe, l'âge, la région d'exercice et le mode d'exercice. Pour les projections de 2008, ont été retenus comme critères la région, le mode d'exercice et le fait d'être généraliste ou spécialiste. En effet, la région de diplôme et le délai à l'entrée dans la vie active ne sont pas des variables d'intérêt pour les résultats des projections. La région d'exercice est nettement corrélée à la zone d'exercice d'une part et d'autre part, les résultats des projections pourraient être analysés au niveau de la région et de la zone d'exercice. La corrélation entre la spécialité et la zone d'exercice choisie n'est pas négligeable, mais d'une part, seules certaines spécialités se distinguent des autres, et d'autre part, il n'est pas possible de retenir à la fois le critère région d'exercice et le critère spécialité. La région d'exercice a donc été privilégiée. Il existe toutefois une différence nette entre généralistes et spécialistes. C'est pourquoi le fait d'être généraliste ou spécialiste a été retenu comme critère. Le mode et la zone d'exercice choisis sont liés. Le mode d'exercice choisi a donc été retenu comme critère. En revanche, la corrélation entre la zone d'exercice choisie et le sexe d'une part et l'âge à l'entrée dans la vie active d'autre part n'ont pas été jugées suffisantes pour que ces variables soient retenues comme critères.

Comme pour l'attribution du mode d'exercice, le choix des critères région d'exercice, généraliste/spécialiste et mode d'exercice conduit à la constitution de strates de petite taille. Cependant, les résultats produits par le tirage aléatoire ne sont pas très éloignées des probabilités-cible.

Le modèle n'attribue pas de zone d'exercice aux médecins choisissant comme région d'exercice à l'entrée dans la vie active un DOM, un TOM ou la Corse (voir A. 1. a.). Les probabilités de choisir

une zone d'exercice donnée sont donc calculées uniquement pour les médecins commençant à exercer dans les régions autres que les DOM-TOM et la Corse.

Les probabilités sont supposées égales à celles observées au cours des dernières années et maintenues constantes pendant toute la durée de la période de projection, sauf si le scénario simulé ne retient pas l'hypothèse de comportements constants en ce qui concerne le choix de la zone d'exercice à l'entrée dans la vie active.

Le modèle attribue une zone d'exercice par un tirage aléatoire (voir A. 1. d.). Dans le cas où il doit attribuer une zone d'exercice à un médecin appartenant à une strate pour laquelle la probabilité n'a pas été calculée (ce qui se produit lorsqu'aucun médecin appartenant à cette strate n'est présent dans les fichiers servant à estimer les probabilités), le modèle attribue la zone « pôle urbain avec CHU » (modalité la plus fréquente), sauf si la région attribuée au médecin en question est un DOM, un TOM ou la Corse.

L'information est ensuite ajoutée dans la base des médecins à faire entrer dans la vie active au cours de l'année.

À l'issue du passage des sous-modules 2.1 (attribution d'un délai à l'entrée dans la vie active), 2.2 (attribution d'un sexe et d'un âge à l'entrée dans la vie active aux entrants de l'année), 2.3 (attribution d'une région), 2.4 (attribution d'un mode d'exercice) et 2.5 (attribution d'une zone d'exercice), le modèle a créé une base contenant des informations sur chacun des médecins à faire entrer dans la vie active au cours de l'année courante.

Sont conservées les variables suivantes : année de diplôme, région de diplôme, spécialité, généraliste/spécialiste, année d'entrée dans la vie active, sexe, âge, région, mode et zone d'exercice.

Le contenu de cette base sera ajouté à la base contenant l'ensemble des informations sur les médecins en activité, mais cette opération n'est pas effectuée immédiatement : auparavant, les médecins sortant temporairement ou définitivement de la vie active sont sélectionnés et ôtés de la base et les changements de région, de mode et de zone d'exercice sont simulés (voir A. 4., D. et E.). Le modèle met donc la base contenant les informations sur les entrants de l'année en attente, après avoir recalculé l'année de naissance, la seule information de la base des médecins en activité qui n'est pas encore incluse, à ce stade, dans la base des entrants de l'année.

## ***6. Les simulations possibles pour le module 2***

Il est possible de modifier le taux de fuites, c'est-à-dire la proportion d'une promotion de diplômés qui n'exercera jamais la médecine en France, ainsi que :

- les probabilités d'entrer dans la vie active l'année de l'obtention du diplôme, un an après ou deux ans ou plus après
- la répartition des nouveaux diplômés entrant dans la vie active par sexe et par âge
- la répartition des nouveaux diplômés entrant dans la vie active par région d'exercice
- la répartition des nouveaux diplômés entrant dans la vie active par mode d'exercice
- la répartition des nouveaux diplômés entrant dans la vie active par zone d'exercice

Concernant ces probabilités, sont paramétrables le nombre d'années récentes sur lesquelles sont calculées celles correspondant à l'hypothèse de comportements constants ainsi que leur valeur (qui peut être modifiée pour faire d'autres hypothèses que celle de comportements constants).

## **D – Module 3 – Les changements au cours de la vie active**

Ce module simule les changements qui interviennent dans la situation professionnelle des médecins au cours de leur vie active, c'est-à-dire les changements de région d'exercice, de mode d'exercice, de zone d'exercice, les cessations temporaires d'activité et les reprises d'activité.

Ce module concerne uniquement les médecins en activité au cours de l'année  $n$  qui restent actifs au cours de l'année  $n+1$ . Les probabilités sont donc estimées, à partir des données observées, sur le champ des médecins actifs au cours de deux années consécutives et s'appliquent, pour l'année  $n$  de la période de projection, aux médecins qui seront encore actifs en  $n+1$ , ce qui signifie que la simulation des changements de région, de mode et de zone d'exercice sera effectuée une fois les médecins cessant temporairement ou définitivement leur activité ôtés de la base des médecins en activité et avant que les médecins reprenant une activité ou entrant dans la vie active ne soient intégrés ou réintégrés à la base (voir A. 4.).

Ces événements n'étaient pas modélisés dans la version précédente du modèle.

### ***1. Sous-module 3.1 : le changement de région d'exercice***

#### ***a. Le choix de la méthode***

Les médecins changeant de région d'exercice au cours d'une année sont assez peu nombreux au regard de l'effectif de médecins actifs. Entre 2005 et 2006 par exemple, 1,4 % des médecins actifs en 2005 et 2006 ont changé de région d'exercice.

Il est possible de calculer des probabilités pour des médecins exerçant dans une région  $r$  au cours de l'année  $n$  d'exercer dans une autre région  $r'$  en  $n+1$ . On obtient alors une matrice de probabilités de dimensions  $\text{nbr} \times \text{nbr}$  où  $\text{nbr}$  est le nombre de régions. Il est également possible de décliner ces probabilités selon un certain nombre de critères choisis parmi les informations relatives aux médecins utilisées dans le cadre des projections (sexe, âge, spécialité, mode, région et zone d'exercice notamment). Cependant, les probabilités ainsi obtenues (par exemple, la probabilité pour un médecin exerçant en  $n$  en Picardie d'exercer en Lorraine en  $n+1$  ou - dans le cas où d'autres critères que la région d'exercice en  $n$  sont retenus - la probabilité pour une femme généraliste exerçant en Picardie en  $n$  d'exercer en Lorraine en  $n+1$ ) seraient extrêmement faibles, ce qui rendrait particulièrement variables les résultats des tirages à l'aide desquels serait réalisée la simulation, ces tirages étant d'autant plus aléatoires que les probabilités-cibles sont faibles (voir communication aux Journées de Méthodologie Statistique 2009).

Par conséquent, nous avons opté pour une autre méthode, consistant à sélectionner dans un premier temps, pour une année donnée de la période de projection, les médecins changeant de région entre  $n$  et  $n+1$ , puis dans un second temps, à attribuer une nouvelle région d'exercice aux médecins sélectionnés au cours de la première étape pour changer de région. L'événement « rare » (quitter la région A pour la région B) est ainsi décomposé en deux événements moins « rares » : changer de région et, pour les médecins changeant de région, choisir la région B comme nouvelle région d'exercice.

#### ***b. L'estimation des probabilités et la simulation des changements de région***

Sont donc estimées des probabilités de changer de région au cours d'une année donnée, puis, pour les médecins changeant de région au cours d'une année donnée, des probabilités de choisir chacune des régions comme nouvelle région d'exercice. Ces deux jeux de probabilités sont estimés à partir des fichiers de l'Ordre, appariés deux à deux.

Les probabilités de changer de région sont estimées sur le champ des médecins déclarés actifs (réguliers ou non) dans deux fichiers annuels successifs, puisqu'elles seront appliquées à ce même champ au cours de la simulation. Les probabilités de choisir une région donnée de destination sont quant à elles calculées sur le champ des médecins déclarés actifs dans les deux fichiers annuels successifs et ayant changé de région entre les deux années consécutives observées.

La proportion de médecins ayant changé de région, de même que la répartition des médecins ayant changé de région entre deux années consécutives par région de destination sont calculées pour un certain nombre d'années, paramétrable. Le nombre d'années maximal n'est pas fixé, mais il ne peut être supérieur au nombre de fichiers annuels de l'Ordre disponibles, moins un (si l'on dispose de 4 fichiers, il est possible de faire 3 appariements). En 2008, les 4 derniers fichiers (soit 3 appariements) ont été utilisés pour estimer ces deux jeux de probabilités.

Les probabilités utilisées pour la simulation correspondent aux proportions ou répartitions moyennes.

Les probabilités sont déclinées selon les critères retenus parmi les suivants : la spécialité, le fait d'être généraliste ou spécialiste, le sexe, l'âge (en tranches quinquennales), la région d'exercice, le mode d'exercice, la zone d'exercice.

Pour les probabilités de changer de région, nous avons retenu pour les projections de 2008 comme critères la région d'exercice, l'âge, le mode d'exercice et le fait d'être généraliste ou spécialiste. La région de départ est liée à la région d'arrivée, les flux se produisant assez souvent entre régions géographiquement proches. De plus, la probabilité de changer de région est corrélée à la région d'exercice. Il est donc important de prendre en compte la région de départ pour sélectionner les médecins qui changeront de région, afin d'estimer correctement le flux de départs dans chacune des régions, qui détermine, avec le flux d'entrées, le solde migratoire de la région, et finalement les effectifs régionaux de médecins actifs. Ce critère a été retenu. L'âge est nettement corrélé au fait de changer ou non de région, de même que le mode d'exercice. Ces deux critères ont donc également été retenus. La corrélation entre la spécialité et le fait de changer ou non de région n'est pas nulle, mais elle n'est pas stable dans le temps. Ce critère a donc été écarté et le fait d'être généraliste ou spécialiste a été retenu, bien que la corrélation entre cette variable et le fait de changer ou non de région soit modérée. Le sexe n'est pas très corrélé au fait de changer ou non de région et n'a donc pas été sélectionné. Le lien entre la zone d'exercice et le changement de région est plutôt net, mais du fait du moindre intérêt porté aux résultats des projections par zone d'exercice par rapport aux résultats par mode d'exercice, par région ou selon le fait d'être généraliste ou spécialiste, nous avons privilégié les autres variables déjà retenues comme critères.

Pour les probabilités de choisir une région  $r$  comme nouvelle région d'exercice, nous avons retenu uniquement la région d'exercice de départ (avant le changement), nettement corrélée à la région de destination. Il n'est pas possible de prendre en compte d'autres critères, étant donné la petite taille des strates obtenues avec le seul critère région.

Les probabilités de changer de région et les probabilités de choisir une région donnée comme nouvelle région d'exercice sont supposées égales à celles observées au cours des années récentes, et restent constantes tout au long de la période de projection, sauf si une variante repose sur une hypothèse autre que celle de comportements constants en matière de changements de région.

L'attribution d'une nouvelle région d'exercice est réalisée en deux temps. Dans un premier temps, les médecins changeant de région entre l'année  $n$  et l'année  $n+1$  sont sélectionnés par un tirage aléatoire (voir A. 1. d.). Dans un second temps, un autre tirage aléatoire attribue une nouvelle région d'exercice aux médecins sélectionnés pour changer de région.

Si la probabilité de changer de région n'a pas été calculée pour une strate donnée, aucun médecin correspondant à cette strate n'ayant été observé dans les données ayant servi à calculer les probabilités, la probabilité utilisée par le modèle est nulle. Si le modèle doit attribuer une nouvelle région d'exercice à un médecin appartenant à une strate pour laquelle la probabilité n'a pas été calculée, il attribue la région PACA aux médecins quittant l'Ile-de-France et la région Ile-de-France aux autres médecins.

La région d'exercice n'est pas mise à jour immédiatement à l'issue de cette étape du programme. En effet, dans le cas où la région est utilisée comme critère pour l'une des deux étapes suivantes (changement de mode d'exercice ou de zone d'exercice), les calculs doivent être effectués sur la région d'exercice de départ, et non sur la région d'arrivée.

Les mises à jour des variables région, mode et zone d'exercice des médecins vivant un changement à l'issue de l'année courante de la période de projection sur l'une de ces trois caractéristiques sont effectués simultanément, une fois simulés les changements de région, de mode et de zone d'exercice.

## ***2. Sous-module 3.2 : le changement de mode d'exercice***

La proportion de médecins changeant de mode d'exercice entre deux années consécutives est également très faible. Pour simuler les changements de mode d'exercice, on procède donc comme pour simuler les changements de région : on sélectionne d'abord, par un premier tirage aléatoire, les médecins qui changent de mode d'exercice, puis, par un second tirage aléatoire, on attribue un nouveau mode d'exercice aux médecins changeant de mode d'exercice.

Les probabilités de changer de mode d'exercice et de choisir tel ou tel nouveau mode d'exercice sont estimées à partir des fichiers de données annuelles de l'Ordre, appariés deux à deux. Ces probabilités sont calculées sur le champ auquel elles seront appliquées : les probabilités de changer de mode d'exercice sont estimées sur le champ des médecins en activité (en activité régulière ou remplaçants) au cours de deux années consécutives et les probabilités de choisir tel ou tel nouveau mode d'exercice sont calculées sur le champ des médecins en activité au cours de deux années consécutives et ayant changé de mode d'exercice.

Les deux jeux de probabilités estimées correspondent à la moyenne des proportions et des répartitions calculées pour un certain nombre d'années, paramétrable. En 2008, trois appariements ont été utilisés (2005/2006, 2004/2005 et 2003/2004) pour calculer les répartitions des médecins actifs au cours de deux années consécutives selon qu'ils ont ou non changé de mode d'exercice entre l'année  $n$  et l'année  $n+1$  et les répartitions des médecins actifs au cours de deux années consécutives et ayant changé de mode d'exercice selon leur nouveau mode d'exercice. Le nombre d'années utilisées n'est limité que par le nombre de fichiers disponibles. Toutefois, il est préférable de ne pas estimer les probabilités sur des données trop anciennes, qui correspondraient à des comportements peut-être trop différents des comportements futurs.

Pour le calcul des probabilités, on pourra retenir comme critères : la spécialité, le fait d'être généraliste ou spécialiste, le sexe, l'âge (en tranches quinquennales), la région d'exercice, le mode d'exercice, la zone d'exercice, et le fait de changer de région d'exercice (en même temps que de mode d'exercice). Pour les projections de 2008, les critères retenus pour le calcul des probabilités de changer de mode d'exercice sont le mode d'exercice, l'âge et la spécialité (voir encadré 3). Pour le calcul des probabilités de choisir tel ou tel nouveau mode d'exercice, les résultats de plusieurs modélisations ont conduit à retenir comme critères le mode d'exercice, l'âge, le fait d'être généraliste ou spécialiste et le sexe.

Les probabilités sont supposées égales à celles observées au cours des années récentes et constantes tout au long de la période de projection, pour tous les scénarios retenant l'hypothèse de comportements constants en matière de changement de mode d'exercice.

L'attribution d'un nouveau mode d'exercice aux médecins changeant de mode d'exercice est réalisée en deux temps. Les médecins changeant de mode d'exercice entre l'année  $n$  et l'année  $n+1$  sont sélectionnés par un premier tirage aléatoire (voir A. 1. d.). Ensuite, un nouveau mode d'exercice est attribué aux médecins sélectionnés pour changer de mode d'exercice par un second tirage aléatoire.

Lorsque la probabilité de changer n'a pas été calculée, elle est supposée nulle. Lorsque le modèle doit attribuer un nouveau mode d'exercice à un médecin changeant de mode d'exercice et appartenant à une strate pour laquelle les probabilités n'ont pas été calculées, il attribue le mode libéral aux médecins quittant un autre mode d'exercice, et le mode « mixte » aux médecins quittant le mode libéral.

La mise à jour de la variable mode d'exercice n'est pas effectuée immédiatement, afin que les calculs relatifs à la simulation du changement de zone d'exercice puissent être effectués à partir de l'ancien mode d'exercice des médecins ayant changé de mode d'exercice, si la variable indiquant si le médecin a changé de mode d'exercice est retenue comme critère pour cette étape.

### ***3. Sous-module 3.3 : le changement de zone d'exercice***

La proportion de médecins changeant de zone d'une année à la suivante est également faible. La simulation des changements de zone d'exercice est donc effectuée en deux temps, comme la simulation des changements de région et de mode d'exercice.

Pour le changement de zone d'exercice, plusieurs cas se présentent : les médecins peuvent changer de zone d'exercice tout en ne quittant pas la métropole, en changeant de région ou non ; ils peuvent quitter un DOM, un TOM ou la Corse pour venir en métropole ; ils peuvent changer de zone d'exercice parce qu'ils quittent la métropole pour un DOM, un TOM ou la Corse. Rappelons que lorsqu'un médecin exerce dans un DOM, un TOM ou en Corse, il ne lui est pas attribué de zone d'exercice. La variable zone d'exercice prend alors une modalité spécifique (zonex = 5 : DOM-TOM ou Corse). Pour les médecins qui quittent la métropole pour un DOM, un TOM ou la Corse, la variable zonex prend donc systématiquement la valeur 5 (DOM-TOM ou Corse). Si un médecin reste dans l'ensemble DOM-TOM ou Corse, il ne change pas de zone, même s'il est mobile au sein de cet ensemble.

Le modèle traite séparément le cas où un médecin reste en métropole, qu'il change ou non de région, et le cas où il quitte un DOM, un TOM ou la Corse pour la métropole.

#### ***a. Le médecin reste en métropole***

Le modèle estime les probabilités de changer de zone d'exercice pour les médecins en activité (activité régulière ou remplaçants) dans une région métropolitaine au cours de deux années consécutives, qu'il s'agisse de la même région ou non. Il calcule également les probabilités pour un médecin en activité dans une région métropolitaine au cours de deux années consécutives et ayant changé de zone d'exercice, de choisir telle ou telle nouvelle zone d'exercice. Ces deux jeux de probabilités sont estimés à partir des fichiers de données annuels de l'Ordre, appariés deux à deux.

Les probabilités de changer de zone d'exercice sont calculées sur le champ des médecins actifs et exerçant en métropole au cours de deux années consécutives. Les probabilités de choisir telle ou telle nouvelle zone d'exercice sont calculées sur celui des médecins actifs et exerçant en métropole au cours de deux années consécutives, ayant changé de zone entre les deux années.



Le modèle calcule la répartition des médecins selon qu'ils ont ou non changé de zone d'une part et la répartition des médecins ayant changé de zone selon la nouvelle zone d'exercice d'autre part pour un nombre d'années paramétrable. Le nombre d'années utilisables n'est limité que par le nombre de fichiers de l'Ordre disponibles. En 2008, 4 fichiers annuels (donc 3 appariements) ont été utilisés : 2003 à 2006. Les probabilités estimées et appliquées aux médecins actifs au cours des simulations correspondent à la moyenne des proportions et des répartitions calculées.

Les probabilités sont déclinées selon les critères retenus parmi la spécialité, le fait d'être généraliste ou spécialiste, le sexe, l'âge (en tranches quinquennales), la région d'exercice, le mode d'exercice, la zone d'exercice, le fait de changer de région d'exercice et le fait de changer de mode d'exercice (en même temps que de zone d'exercice). En 2008, pour les probabilités de changer de zone d'exercice, les critères zone d'exercice, âge, mode d'exercice et le fait d'être généraliste ou spécialiste ont été retenus. La zone d'exercice, l'âge et le mode d'exercice sont nettement corrélés au fait de changer ou non de zone d'exercice. Le fait d'être généraliste ou spécialiste l'est moins, mais il importe de ne pas déformer à tort la répartition des généralistes et des spécialistes par zone d'exercice (les différentes spécialités se distinguent peu les unes des autres). La corrélation entre le sexe et le fait de changer ou non de zone d'exercice n'est pas nettement significative, et le lien entre la région d'exercice et le fait de changer ou non de zone n'est pas stable dans le temps. Pour les probabilités de choisir telle ou telle nouvelle zone d'exercice, la zone d'exercice, le fait d'être généraliste ou spécialiste et l'âge ont été retenus. Le choix de la zone d'exercice comme critère est motivé par la corrélation entre la zone d'exercice de départ et la zone d'exercice d'arrivée, mais aussi par le souci de ne pas déformer à tort la répartition des médecins par zone d'exercice. C'est aussi essentiellement pour ne pas déformer à tort la répartition des généralistes et des spécialistes par zone d'exercice que nous avons retenu comme critère le fait d'être généraliste ou spécialiste, la corrélation n'étant pas très forte. L'âge est nettement corrélé au choix de la nouvelle zone d'exercice, contrairement au sexe et au mode d'exercice, qui le sont moins et qui n'ont par conséquent pas été retenus. Retenir la région d'exercice (assez corrélée à la zone d'exercice) comme critère était difficile, compte tenu du nombre de modalités de cette variable.

Les probabilités de changer de zone d'exercice et les probabilités de choisir telle ou telle zone d'exercice sont supposées égales à celles observées au cours de la période récente et constantes pendant toute la période de projection, sauf si l'on ne retient pas, pour un scénario donné, l'hypothèse de comportements constants en ce qui concerne le changement de zone d'exercice.

Les changements de zone d'exercice sont simulés en deux temps : un premier tirage aléatoire (voir A. 1. d) sélectionne les médecins changeant de zone d'exercice, en fonction des probabilités estimées, uniquement parmi les médecins que la simulation a fait rester en métropole au moment de la simulation des changements de région. Un second tirage aléatoire attribue une nouvelle zone d'exercice aux médecins sélectionnés pour changer de zone d'exercice, selon les probabilités estimées.

Si, pour une strate donnée, la probabilité de changer de zone n'a pas été calculée, aucun médecin n'ayant les caractéristiques correspondant à cette strate dans les données ayant servi à l'estimation des probabilités, la probabilité de changer de zone est supposée nulle. Lorsque le modèle doit attribuer une nouvelle zone d'exercice à un médecin appartenant à une strate pour laquelle les probabilités n'ont pas été calculées, il lui attribue la zone 1 (pôle urbain avec CHU), sauf s'il quitte cette zone, auquel cas il lui attribue la zone 2 (pôle urbain sans CHU).

### ***b. Le médecin quitte un DOM, un TOM ou la Corse, pour la métropole***

Il n'est pas nécessaire de sélectionner dans un premier temps les médecins changeant de zone parmi ces médecins, qui changent tous de zone puisque la valeur de la variable zone d'exercice est spécifique pour les médecins exerçant dans un DOM, un TOM ou la Corse.

Les probabilités de choisir une zone d'exercice donnée appliquées aux médecins quittant un DOM, un TOM ou la Corse sont estimées de la même façon que les probabilités de changer de zone d'exercice pour les médecins restant en métropole, mais sont calculées sur un champ différent : celui des médecins exerçant dans un DOM, un TOM ou la Corse au cours d'une année donnée et en métropole au cours de l'année suivante. Ces probabilités sont déclinées selon les mêmes critères que les mêmes probabilités estimées pour les médecins restant en métropole. Sauf si la région a été retenue comme critère pour le calcul de ces probabilités, les probabilités ne sont pas déclinées pour chaque région des DOM-TOM ou la Corse, les effectifs étant trop faibles d'une part et les différences entre régions peu marquées.

La mise à jour de la variable zone d'exercice est effectuée à l'issue de cette étape, en même temps que la mise à jour des variables mode d'exercice et région d'exercice.

#### ***4. Sous-module 3.4 : la cessation temporaire d'activité***

Au début de la simulation, est créée une table contenant un enregistrement pour chaque médecin en cessation temporaire d'activité au 31 décembre de l'année précédent le début de la période de projection. Pour les projections de 2008, cette table a été créée à partir du fichier de l'Ordre au 31 décembre 2006. Elle évolue au fur et à mesure de la simulation : les médecins en cessation d'activité vieillissent, décèdent pour certains (voir module 4), reprennent une activité pour d'autres, tandis que de nouveaux médecins cessent temporairement leur activité.

Au cours de chaque année de la période de projection, sont simulées des cessations temporaires d'activité : les enregistrements correspondant aux médecins que le modèle fait cesser temporairement leur activité sont ôtés de la base des médecins en activité et placés dans la base des médecins en cessation temporaire d'activité.

Les probabilités de cesser temporairement son activité médicale sont estimées à partir des fichiers annuels l'Ordre des médecins, appariés deux à deux.

Le champ retenu pour le calcul des probabilités est celui des médecins en activité au cours d'une année  $n$  (en activité régulière ou remplaçants). Les médecins qui cessent temporairement leur activité sont identifiés par une valeur spécifique de la variable indiquant la nature de leur activité dans le fichier de l'Ordre de l'année  $n+1$ .

Pour chaque appariement de fichiers, est calculée la répartition des médecins actifs selon qu'ils ont cessé temporairement leur activité ou non entre les deux années observées. Cette répartition est calculée pour chaque appariement utilisé pour le calcul des probabilités. Le nombre d'appariements utilisables n'est limité que par le nombre de fichiers de l'Ordre disponibles, mais il est préférable de ne pas utiliser des fichiers trop anciens, afin d'estimer les probabilités à partir de comportements observés assez récents. En 2008, 3 appariements (c'est-à-dire les 4 fichiers) ont été utilisés. Les probabilités estimées correspondent à la moyenne des répartitions calculées.

Les probabilités de cesser temporairement son activité sont déclinées selon les critères retenus. Il est ainsi possible de les décliner selon la spécialité, le fait d'être généraliste ou spécialiste, le sexe, l'âge (en tranches quinquennales), la région d'exercice, le mode d'exercice, la zone d'exercice. Pour les projections de 2008, nous avons retenu comme critères le sexe, l'âge, le fait d'être généraliste ou spécialiste et le mode d'exercice. L'âge, le sexe et le fait d'être généraliste ou spécialiste sont nettement corrélés au fait de cesser temporairement ou non son activité (les spécialités se distinguent peu les unes des autres en revanche). La corrélation avec le mode d'exercice est quant à elle nette, mais un peu moins forte et un peu plus instable dans le temps, mais il importe de ne pas déformer la répartition des médecins selon le mode d'exercice. La corrélation entre la région d'exercice et le fait

de cesser temporairement ou non son activité n'est pas très forte, et instable dans le temps, tandis que celle entre la zone d'exercice et le fait de cesser ou non temporairement son activité est faible.

Si le modèle doit sélectionner les médecins cessant temporairement leur activité au sein d'une strate pour laquelle la probabilité de cesser temporairement son activité n'a pas été calculée, aucun médecin appartenant à cette strate n'ayant été observé dans les données ayant servi à l'estimation des probabilités, la probabilité de cesser son activité est supposée nulle.

Les cessations temporaires d'activité sont simulées uniquement pour les médecins de moins de 80 ans. A 80 ans, les médecins partent tous à la retraite dans la simulation, sauf s'ils décèdent. Les médecins cessant temporairement leur activité sont sélectionnés en même temps que les médecins partant à la retraite et que les médecins décédant (voir A. 4.).

Une fois les médecins cessant temporairement leur activité sélectionnés par le modèle, les enregistrements correspondant sont supprimés de la base des médecins en activité. Les enregistrements correspondant aux nouveaux médecins ayant cessé leur activité seront ajoutés à la table des médecins en cessation temporaire d'activité, table dans laquelle sont sélectionnés également par le modèle les médecins reprenant une activité. Mais cet ajout est effectué plus tard, après la sélection des médecins reprenant une activité au cours de l'année courante de la période de projection (voir A. 4.).

## **5. Sous-module 3.5 : la reprise d'activité**

### **a. La sélection des médecins reprenant une activité**

Les médecins reprenant une activité sont sélectionnés parmi les médecins en cessation temporaire d'activité : le modèle sélectionne donc certains enregistrements dans la table contenant les enregistrements correspondant aux médecins en cessation temporaire d'activité, les ôte de cette table et les place dans la base des médecins en activité.

Les probabilités de reprendre une activité sont appliquées aux médecins en cessation d'activité. Ces probabilités sont estimées à partir des fichiers annuels de l'Ordre appariés deux à deux. En 2008, 3 appariements (c'est-à-dire les 4 fichiers) ont été utilisés. Le nombre d'appariements utilisés, paramétrable, n'est limité que par le nombre de fichiers de l'Ordre disponibles. Il faut cependant veiller à utiliser des données récentes pour observer des comportements susceptibles de ne pas être trop éloignés des comportements futurs. Les probabilités estimées correspondent à la moyenne des répartitions des médecins en cessation temporaire d'activité au cours d'une année donnée selon qu'ils ont ou non repris une activité au cours de l'année suivante.

Le champ retenu pour le calcul des probabilités de reprendre une activité est celui des médecins en cessation temporaire d'activité. Les médecins reprenant une activité entre l'année  $n$  et l'année  $n+1$ , sont en cessation temporaire d'activité au cours de l'année  $n$  et en activité régulière ou remplaçants au cours de l'année  $n+1$ .

Les probabilités sont déclinées selon les critères retenus parmi le sexe, l'âge, la spécialité et le fait d'être généraliste ou spécialiste. Les médecins en cessation temporaire d'activité n'ont, par définition, ni région, ni mode, ni zone d'exercice. Les variables région, mode et zone d'exercice ne peuvent donc pas être retenues comme critères. Pour l'exercice de projections de 2008, seul l'âge a été retenu comme critère. C'est la seule variable qui soit significativement corrélée à la probabilité de reprendre une activité.

Le modèle sélectionne les médecins reprenant une activité parmi les médecins en cessation d'activité en même temps qu'il sélectionne les médecins décédant ou partant à la retraite parmi les médecins en cessation d'activité (voir module 4 et A. 4.). Seuls les médecins âgés de moins de 80 ans peuvent reprendre une activité. Les médecins en cessation temporaire âgés de 80 ans partent tous à la retraite, sauf s'ils décèdent. La sélection est effectuée à l'aide d'un tirage aléatoire (voir A. 1. d.).

Si le modèle doit sélectionner les médecins reprenant une activité au sein d'une strate pour laquelle la probabilité de reprendre une activité n'a pas été calculée, aucun médecin appartenant à cette strate n'ayant été observé dans les données ayant servi à l'estimation, cette probabilité est supposée nulle.

### ***b. L'attribution d'un mode, d'une région et d'une zone d'exercice aux médecins reprenant une activité***

Il faut ensuite attribuer aux médecins reprenant une activité une région, un mode et une zone d'exercice. En effet, les médecins ne restent pas forcément dans la même région et dans la même zone et ne gardent pas forcément le même mode d'exercice lorsqu'ils reprennent une activité.

On estime donc des probabilités de choisir telle ou telle région, des probabilités de choisir telle ou telle zone d'exercice et des probabilités de choisir tel ou tel mode d'exercice pour les médecins reprenant une activité.

Ces probabilités sont estimées à partir des fichiers annuels de l'Ordre, appariés deux à deux. Le nombre d'appariements utilisés est le même que le nombre d'appariements utilisés pour le calcul des probabilités de reprendre une activité. Le champ est celui des médecins reprenant une activité, c'est-à-dire en cessation temporaire d'activité au cours d'une année donnée et en activité régulière ou remplaçants au cours de l'année suivante.

Les probabilités sont déclinées selon les critères retenus parmi le sexe, l'âge, la spécialité et le fait d'être généraliste ou spécialiste pour l'attribution du mode d'exercice. Le mode d'exercice étant attribué avant la région d'exercice, le mode d'exercice peut être utilisé comme critère pour l'attribution de la région d'exercice. De même, le mode et la région d'exercice peuvent être retenus comme critères pour l'attribution de la zone d'exercice. Pour les projections de 2008, ont été retenus comme critères l'âge, le sexe et le fait d'être généraliste ou spécialiste pour l'attribution du mode d'exercice aux médecins reprenant une activité, le fait d'être généraliste ou spécialiste et le mode d'exercice pour l'attribution de la région d'exercice, et la région d'exercice, le fait d'être généraliste et le mode d'exercice pour l'attribution de la zone d'exercice. Nous avons en fait sélectionné autant que possible les mêmes critères que ceux utilisés pour l'attribution de ces caractéristiques aux médecins entrant dans la vie active ou changeant de région, de mode ou de zone d'exercice.

Après avoir sélectionné les médecins reprenant une activité au cours de l'année, le modèle constitue une base contenant les enregistrements correspondant à ces médecins. Il complète ensuite cette base en attribuant à chaque médecin, successivement, un mode, une région et une zone d'exercice, par trois tirages aléatoires (voir A. 1. d.), en fonction des probabilités estimées.

Si le modèle doit attribuer un mode d'exercice à un médecin reprenant une activité et appartenant à une strate pour laquelle la probabilité de choisir tel ou tel mode d'exercice n'a pas été calculée, il attribue le mode libéral aux généralistes et le mode salarié hospitalier aux spécialistes. Si la probabilité de choisir telle ou telle région n'a pas été calculée pour une strate donnée, il attribue l'Ile-de-France comme région d'exercice. Enfin, si la probabilité de choisir telle ou telle zone d'exercice n'a pas été estimée, il attribue la zone « pôle urbain avec CHU », sauf dans le cas où la région d'exercice (attribuée précédemment) est un DOM, un TOM ou la Corse, auquel cas il attribue la zone d'exercice « DOM-TOM ou Corse ».

### ***c. Fin de la simulation des reprises d'activité***

Une fois la base contenant les enregistrements correspondant aux médecins reprenant une activité complétée (celle-ci contient alors les mêmes informations que la base des médecins en activité), les enregistrements correspondant aux médecins reprenant une activité au cours de l'année courante de l'année de projection sont supprimés de la base des médecins en cessation temporaire, en même temps que les médecins en cessation temporaire qui décèdent ou qui partent à la retraite (voir A. 4.).

Ensuite, les médecins reprenant une activité sont ajoutés à la base des médecins en activité, en même temps que les médecins entrant dans la vie active au cours de l'année courante de la période de projection (voir A. 4.).

## ***6. Les simulations possibles pour le module 3***

Il est possible de modifier :

- les probabilités de changer de région d'exercice
- la répartition des médecins changeant de région par région de destination
- les probabilités de changer de mode d'exercice
- la répartition des médecins changeant de mode par mode de destination
- les probabilités de changer de zone d'exercice
- la répartition des médecins changeant de zone par zone de destination
- les probabilités de cesser temporairement son activité
- les probabilités de reprendre une activité
- la répartition des médecins reprenant une activité par région d'exercice
- la répartition des médecins reprenant une activité par mode d'exercice
- la répartition des médecins reprenant une activité par zone d'exercice

Concernant ces probabilités, sont paramétrables le nombre d'années récentes sur lesquelles sont calculées les probabilités correspondant à l'hypothèse de comportements constants ainsi que la valeur de ces probabilités (que l'on peut modifier pour faire d'autres hypothèses que celle de comportements constants). Le nombre d'années utilisées pour l'estimation des probabilités de changer de région (resp. de mode, de zone) d'exercice et pour l'estimation de la répartition des médecins changeant de région par région (resp. de mode, de zone) de destination est le même. Il en est de même pour les probabilités de reprendre une activité et la répartition des médecins reprenant une activité par région, mode et zone d'exercice.

## **E – Module 4 – La sortie définitive de la vie active**

Le modèle simule deux événements correspondant à une sortie définitive de la vie active : le départ à la retraite et le décès. Les deux événements sont simulés séparément pour permettre l'analyse des flux de médecins sortant de la vie active selon le motif de la cessation définitive d'activité. Peuvent sortir définitivement de la vie active les médecins en activité, mais également les médecins en cessation temporaire d'activité au cours d'une année donnée.

### ***1. Sous-module 4.1 : le départ à la retraite***

Des probabilités de partir à la retraite entre deux années consécutives sont calculées pour les médecins en activité à partir des fichiers individuels de données de l'Ordre, appariés deux à deux. Les mêmes probabilités sont calculées pour les médecins en cessation temporaire d'activité. Les probabilités utilisées pour la simulation correspondent à la moyenne des répartitions des médecins actifs (resp. en cessation temporaire) au cours d'une année  $n$  selon qu'ils sont ou non partis à la retraite en  $n+1$ . Le

nombre de répartitions (c'est-à-dire d'années) utilisées pour le calcul des probabilités est paramétrable et n'est limité que par le nombre de fichiers de l'Ordre disponibles, mais il faut veiller à ne pas utiliser des données trop anciennes. Pour les projections de 2008, ont été utilisées comme probabilités la moyenne des deux dernières répartitions observées, afin de ne pas prendre en compte pour l'estimation des probabilités des enregistrements correspondant à des médecins partis à la retraite dans le cadre du dispositif MICA.

Les probabilités de partir à la retraite qui seront appliquées aux médecins actifs sont calculées sur le champ des médecins actifs au cours de l'année *n*. Celles qui seront appliquées aux médecins en cessation temporaire sont calculées sur le champ des médecins en cessation temporaire au cours de l'année *n*.

Ces probabilités sont déclinées selon les critères retenus. Pour le calcul des probabilités appliquées aux médecins en activité, on peut retenir le sexe, l'âge (en tranches)<sup>17</sup>, le mode d'exercice, le fait d'être généraliste ou spécialiste, la spécialité, la région d'exercice et la zone d'exercice. Pour le calcul des probabilités appliquées aux médecins en cessation, il ne peut retenir que le sexe, l'âge, le fait d'être généraliste ou spécialiste et la spécialité, puisque ces médecins n'ont ni région, ni mode, ni zone d'exercice. Pour les projections de 2008, le sexe, l'âge, le fait d'être généraliste ou spécialiste et le mode d'exercice ont été retenus pour le calcul des probabilités appliquées aux médecins actifs et l'âge, le sexe et le fait d'être généraliste ou spécialiste pour le calcul des probabilités appliquées aux médecins en cessation temporaire d'activité. La corrélation entre la région d'exercice et le départ à la retraite n'est pas nulle, de même que celle entre la zone d'exercice et le départ à la retraite, mais toutes deux sont moins fortes. Les spécialités se distinguent en revanche peu les unes des autres.

Le modèle sélectionne les médecins partant à la retraite parmi les médecins en activité d'une part, et parmi les médecins en cessation d'activité d'autre part, à l'aide de chacun des jeux de probabilités estimés. Les sélections sont effectuées par des tirages aléatoires (voir A. 1. d.).

Un traitement particulier est effectué pour les médecins âgés de 80 ans : on sélectionne d'abord les médecins qui décèdent au cours de l'année courante de la période de projection. Tous les autres médecins de cet âge partent à la retraite. Le champ des projections est en effet constitué par les médecins de moins de 80 ans.

La probabilité de partir à la retraite est supposée nulle pour les médecins âgés de moins de 50 ans. Dans les données utilisées pour l'estimation des probabilités, quelques médecins partent à la retraite avant cet âge, mais il s'agit probablement d'erreurs de codage.

La sélection des médecins partant à la retraite est effectuée en même temps que la sélection des médecins décédant et des médecins cessant temporairement leur activité, afin que les probabilités estimées soient bien appliquées au champ sur lequel elles ont été estimées (voir A. 4.).

Si le modèle doit sélectionner les médecins partant à la retraite au sein d'une strate pour laquelle la probabilité n'a pas été estimée, aucun médecin n'appartenant à cette strate dans les fichiers utilisés pour l'estimation des probabilités, il utilise une probabilité de partir à la retraite qui est supposée nulle.

Enfin, une fois sélectionnés, les médecins partant à la retraite parmi les médecins actifs sont supprimés de la base des médecins actifs, en même temps que ceux de ces médecins qui décèdent ou cessent temporairement leur activité. Ceux qui partent à la retraite parmi les médecins en cessation temporaire sont supprimés de la base des médecins en cessation temporaire, en même temps que ceux de ces médecins qui décèdent.

---

<sup>17</sup> La variable âge n'est pas recodée en tranches quinquennales comme pour les autres événements, mais en modalités ayant un sens pour le départ à la retraite (événement particulièrement probable à certains âges) : 25-50 ans, 50-58 ans, 59-60 ans, 61-62 ans, 63 ans, 64 ans, 65-70 ans, 71-79 ans et 80 ans.

## ***2. Sous-module 4.2 : le décès***

L'information sur le décès des médecins dans les fichiers de données de l'Ordre n'a pas semblé suffisamment fiable pour servir à l'estimation de probabilités de décéder. Nous avons donc utilisé comme probabilités de décéder les tables de mortalité calculées par sexe et par âge, pour la catégorie « Cadres et professions intellectuelles supérieures » par l'Insee en 2005. Ce faisant, nous faisons une approximation, car il est probable que la mortalité des médecins diffère un peu de celle de l'ensemble des individus de cette catégorie socioprofessionnelle. Toutefois, le différentiel de mortalité peut être supposé relativement faible d'une part, et les effectifs concernés (des médecins décédant au cours de leur vie active) sont très faibles au regard des effectifs de médecins actifs. Ces probabilités sont calculées pour les médecins âgés de 20 à 80 ans.

Le modèle sélectionne les médecins qui décèdent au cours de l'année courante de la période de projection, au sein des médecins en activité d'une part, et au sein des médecins en cessation temporaire d'autre part.

Ces sélections sont effectuées par des tirages aléatoires (voir A. 1. d.), en même temps que les sélections des médecins partant à la retraite et des médecins cessant temporairement leur activité.

Si le modèle doit sélectionner les médecins qui décèdent au sein d'une strate pour laquelle la probabilité de décéder n'est pas connue, il applique une probabilité de décéder nulle.

Enfin, une fois les médecins actifs qui décèdent sélectionnés, ils sont ôtés de la base des médecins actifs, en même temps que les médecins qui partent à la retraite ou qui cessent temporairement leur activité. Ceux qui partent à la retraite parmi les médecins en cessation temporaire sont supprimés de la base des médecins en cessation temporaire, en même temps que les médecins qui décèdent.

## ***3. Les simulations possibles pour le module 4***

Il est possible de modifier :

- les probabilités de partir à la retraite pour les médecins actifs
- les probabilités de partir à la retraite pour les médecins en cessation temporaire d'activité
- les probabilités de décéder

Concernant les probabilités de partir à la retraite, sont paramétrables le nombre d'années récentes sur lesquelles sont calculées les probabilités correspondant à l'hypothèse de comportements constants ainsi que la valeur de ces probabilités (que l'on peut modifier pour faire d'autres hypothèses que celle de comportements constants). En revanche, le nombre d'années utilisées pour l'estimation des probabilités de partir à la retraite pour les médecins actifs d'une part et pour les médecins en cessation temporaire d'activité d'autre part est le même.