

# Tables de mortalité pour la France entière et par département, 1806-1906. Tables de migration nette par département, 1806-1906.

Noël BONNEUIL

*Institut national d'études démographiques,  
133, bld Davout, 75980, Paris cedex 20, France,  
et École des hautes études en sciences sociales*

avril 2005

Dans *Transformation of the French demographic landscape*, 1997, Oxford : Clarendon Press, j'ai remis en question la fiabilité des données récapitulatives de décès, en plus de celle des recensements et des naissances. J'ai introduit les statistiques de décès par âge, dont ne s'était pas servi Etienne van de Walle (*The Female Population of France in the Nineteenth Century* (Princeton : Princeton University Press, 1974)).

Un premier travail (chapitre 3) a consisté à corriger de manière comptable les données de la Statistique Générale de la France (SGF), en utilisant toutes les possibilités de vérification offertes par la mention, dans les ouvrages originaux, des sommes partielles en ligne (le département) ou en colonne (l'âge), ou par la publication de tableaux par état civil, ou pour les « deux sexes réunis ». Bien que cette phase préliminaire n'ait rien à voir avec la reconstruction, et encore moins avec le jugement sur la qualité des recensements, cela me permet de partir de données saines sur le plan comptable.

Désormais, toute l'information disponible dans la SGF est incorporée dans la reconstruction de manière à atteindre la cohérence démographique. Celle-ci signifie que toutes les équations démographiques doivent être vérifiées, et elles seules. Les étapes de la méthode sont les suivantes :

1. Dans chaque département  $i$ , l'ensemble des individus d'un âge donné  $x$  à l'instant  $t$  est soumis à deux forces concurrentes entre l'instant  $t$  et l'instant  $t + dt$  : la mort  $\mu(x, t)$  et l'émigration  $\nu(x, t)$  (équation 5.1 page 57). Une technique due à Greville (1948), utilisée par Chiang (1968) ou Brouard (1986), me permet d'intégrer cette équation au prix de l'hypothèse de calcul la moins forte (voir discussion dans Schoen, 1988). J'ai montré qu'il y a également moyen d'incorporer les immigrants. Pour une durée  $h$ , on aboutit à l'expression du quotient de mortalité  ${}_h q_x^i$  entre les âges  $x$  et  $x + h$  en fonction du total de décès  $D_i([x, x + h])$  entre  $t$  et  $t + h$ , et des effectifs  $p_i(x, t)$  au recensement à

$t$  et  $p_i(x+h, t+h)$  à  $t+h$ . Ces nombres sont tirés de mouvements et de recensements de qualité variable (par exemple, les recensements de 1861 et 1866 sont de bonne qualité, tandis que celui de 1872 est de mauvaise qualité). C'est l'équation 5.5 page 62 de Bonneuil (1997). or,  $D_i([x, x+h])$  est inconnu, et les  $p_i(x, t)$  et  $p_i(x+h, t+h)$  sont tirés de recensements de qualité variable.

2. Les totaux de décès pour un groupe d'âges fixe, par exemple [10,14] ans, sont publiés chaque année à partir de 1856. Ces totaux concernent donc plusieurs cohortes. J'ai proposé pages 47 à 52 de Bonneuil (1997) de retrouver les décès de chaque cohorte parmi le total sur un groupe d'âges donné. Pour cela, j'utilise une distribution par âge et par mois au niveau national, publiée assez régulièrement avec les décès par âge au niveau départemental. Ensuite, à l'aide d'une fonction de lissage bi-dimensionnel sur les décès par âge, telle qu'elle conserve le nombre total enregistré de décès pour chaque classe d'âge, chaque décès est affecté à une cohorte précise. Cette technique numérique évite de recourir à des formules toutes faites, comme celle qui consiste à calculer les quotients d'une cohorte à partir des taux de mortalité concernant des individus d'âges différents. Cette dernière approximation peut fonctionner dans des populations fermées sans trop d'irrégularités en termes d'effectifs de générations et de mortalité, mais, pour les départements français au XIX<sup>e</sup> siècle, c'est précisément ce qu'on cherche, rendre compte de populations ouvertes subissant des fluctuations de naissances et de mortalité.
3. Le simple examen des effectifs  $p_i(x, t)$  (figure 2.6 page 36 de van de Walle (1974) repris pages 44 et 45 de Bonneuil (1997)) montre que ceux-ci peuvent être sous-estimés, déformés par l'attraction aux âges ronds, ou même par des mécomptes (notamment au recensement de 1872). Un traitement préalable permet de surmonter ces difficultés. Il fournit des estimations  $p_i(x, t)$  et  $p_i(x+h, t+h)$  qui s'incorporent à l'équation 5.5 page 62 de Bonneuil (1997). On obtient alors les quotients de mortalité  ${}_h q_x^i$ . Cela ne présume pas de l'estimation finale des  $p_i(x, t)$ .
4. Pour ce qui concerne l'attraction aux âges ronds, diverses techniques sont proposées par exemple dans le Manuel X des Nations unies<sup>1</sup>. Il est difficile en général de distinguer entre un effet d'attraction aux âges ronds et le passage d'une génération plus nombreuse. Ici, le fait de disposer de séries de recensements, tous les cinq ans de 1851 à 1906, permet de lever cette ambiguïté : j'ai proposé de repérer préalablement les cohortes au fil des recensements<sup>2</sup> et d'observer les effectifs de cohortes. Ces derniers montrent des fluctuations régulières avec des pics aux âges se terminant par des chiffres ronds et des creux aux autres. Van de Walle avait remarqué ce phénomène et avait exclu un effet de migration. J'ai repris totalement sa discussion et j'ai proposé un

---

<sup>1</sup>Manuel X, Indirect techniques for demographic estimation. New York: United Nations, 1983.

<sup>2</sup>Une première opération d'ajustement est nécessaire parce que la durée n'est pas exactement de 5 ans, mais peut être de 4 ans et demi (décembre 1881 à mai 1886), ou même de 6 ans entre celui de 1866 et celui d'avril-mai 1872.

lissage aux moindres carrés. Les figures 4.1, 4.2 et 4.3 pages 41 à 43, et A1, A2 et A3 pages 174 à 176 de Bonneuil (1997) montrent que ce lissage étête soigneusement les irrégularités dans le cas de la Creuse et du Finistère pris comme exemples. La figure 1 représente le cas du département de la Charente pour les cohortes 1826-30 à 1836-40.

5. Les quotients  ${}_h q_x^i$  entre  $t$  et  $t+h$  sont calculés pour chaque âge (de la naissance à 0-4 ans, de 0-4 ans à 5-9 ans, etc.), chaque période quinquennale, et chaque département indépendamment les uns des autres, à part  $p_i(x, t)$  qui joue un rôle, même s'il est corrigé entre-temps, dans l'estimation de  ${}_h q_x^i([t, t+h])$  et de  ${}_h q_x^i([t-h, t])$ . Un premier test de plausibilité consiste à dresser les tables de mortalité ainsi estimées pour chaque département et chaque période, et à les comparer à des tables-type, de Ledermann par exemple. L'exemple de la Charente est donné sur la figure 2, ceux du Finistère, de la Creuse, de la Gironde et de la Seine sont dans Bonneuil (1997), celui de l'Hérault dans Bonneuil (1998).

Le passage d'une table-type à une table-type en perspectif est présenté en annexe. L'adéquation à une table de Ledermann à simple entrée selon une distance choisie (voir Bonneuil, 1997, p. 63 et 64 pour le choix de la distance) révèle d'une part l'amélioration de la qualité des données à partir du recensement de 1856, d'autre part un décalage à peu près systématique à la naissance et à 0-4 ans. ce décalage est gommé si on utilise des tables à double entrée. Bien que l'ajustement à une table-type à simple entrée soit moins souple, il fournit un moyen simple, sans doute imparfait, d'évaluer le sous-enregistrement, qui se concentre aux jeunes âges. Il peut arriver que le quotient de Ledermann soit inférieur à celui estimé directement. cela arrive certaines années notamment pour la Haute-Vienne. Dans ces cas (largement minoritaires), il faut considérer soit la présence de double comptes, soit que la mortalité infantile est effectivement supérieure à celle de la Ledermann la plus proche, et que le sous-enregistrement des naissances est nul.

Les tables directement estimées ont incorporé des données biaisées (les effectifs  $p_i(x, t)$  et  $D_i([x, x+h])$ ). Aussi un moyen de les redresser et d'atteindre à la cohérence démographique est-il de remplacer ces tables par les tables-type de Ledermann les plus proches au sens de la minimisation de la distance choisie. Cette opération effectuée, on peut estimer à nouveau les  $p_i(x, t)$  à partir des  $p_i(x+h, t+h)$  et des quotients de Ledermann  ${}_h q_x^{i,L}([t, t+h])$  qui ont remplacé les quotients directement estimés  ${}_h q_x^i([t, t+h])$ . Les équations démographiques sont satisfaites, sans ajout de contrainte *ad hoc*, en collant au plus près des données, et sans surestimer la qualité de ces dernières.

6. L'équation 5.1 page 51 de Bonneuil (1997) exprimant l'évolution de l'effectif  $p_i(x, t)$  permet de calculer le quotient de mortalité et le quotient de migration nette (équation 5.7 page 62), par âge, par département et par période quinquennale indépendamment les uns des autres (hormis les effectifs  $p_i(x, t)$  qui jouent un rôle dans deux équations, comme je l'ai déjà mentionné). La

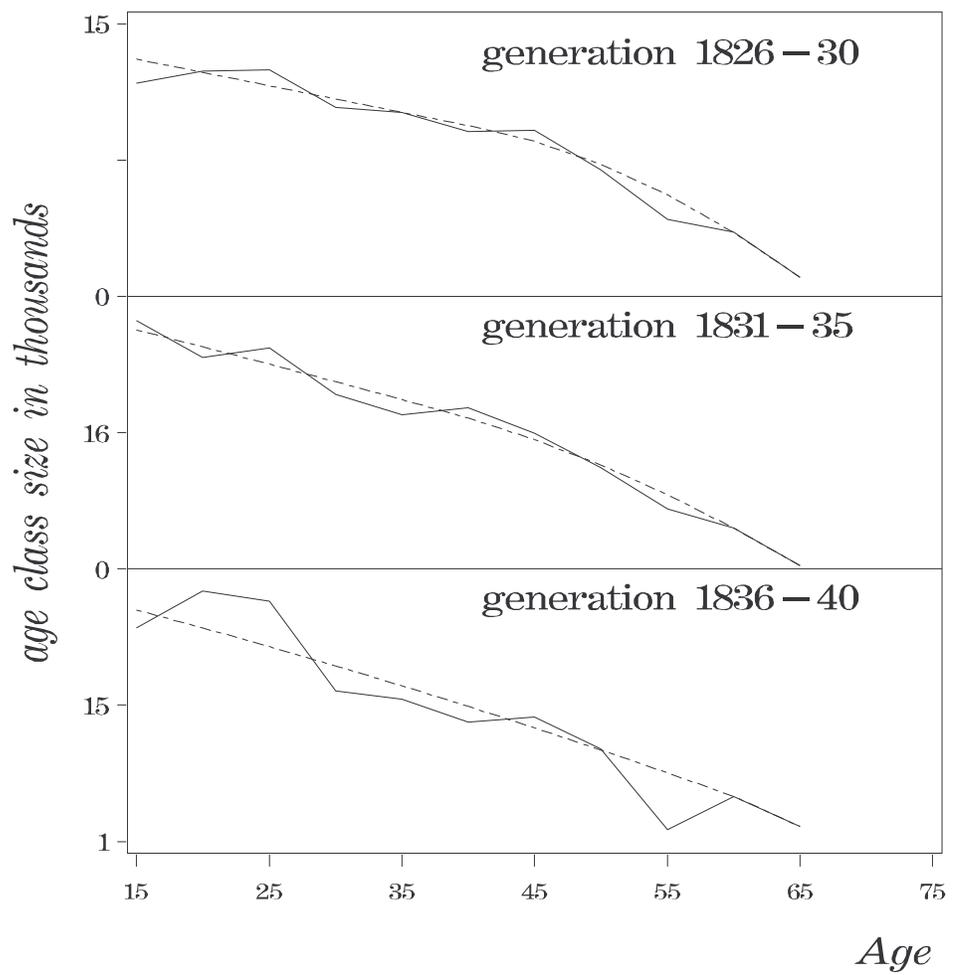


Figure 1: Traitement de l'attraction aux âges ronds par lissage aux moindres carrés par cohorte. Charente, générations 1826-30 à 1836-40

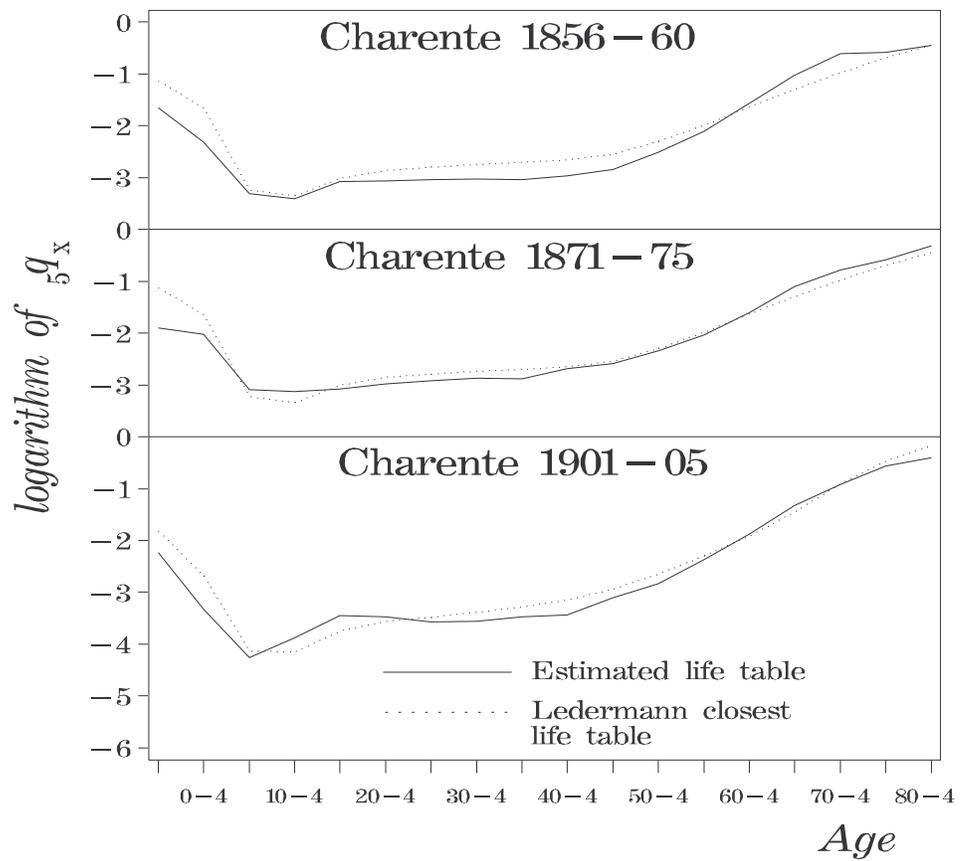


Figure 2: Exemple d'ajustement des quotients de mortalité reconstruits par des quotients de Ledermann. Charente

trajectoire en temps et en âge des différents taux de migration nette montre une grande cohérence par département, ce qui valide a posteriori la reconstruction, de même que la forte adéquation entre les tables de mortalité estimées directement et les tables-type de Ledermann.

7. Les résultats dépendent des techniques utilisées, par exemple du lissage préalable pour se débarrasser de l'attraction aux âges ronds. J'ai présenté pages 67 à 72 une analyse de sensibilité.
8. De 1806 à 1856, les statistiques de décès par âge n'ont pas été publiées. La reconstruction de 1856 à 1906 a fourni non seulement les taux de migration nette par âge, mais aussi un critère de correction des tables-type estimées. En m'appuyant sur ces deux résultats, j'ai mené une reconstruction de 1806 à 1856. Je biaise donc délibérément les tables-type utilisées en « copiant » pour 1806-1856 la tendance du biais observé de 1856 à 1906, et j'ai prolongé en arrière dans le temps les taux de migration nette par âge et par département. En imposant que la pyramide des âges en 1856, obtenue par la reconstruction en avant de 1806 à 1856, soit identique à celle obtenue par la reconstruction en arrière, pas à pas, de 1906 à 1856, il est possible de proposer des corrections à la fois du recensement de 1806 et des naissances de 1806 à 1855 (les autres recensements de cette période ne sont pas utilisables). Il se trouve que la tendance des estimations de sous-enregistrement de 1806 à 1856 est la même, pour chaque département, que celle trouvée de 1856 à 1906, les deux méthodes étant pourtant très différentes. La correction du recensement de 1806 a pour effet de relever l'estimation de la population féminine de 1806 à 1851 (15,2 millions en 1806 contre 15,0 estimés par van de Walle, 18,5 millions en 1851 contre 18,3 estimés par van de Walle pour la France entière).
9. Les sous-enregistrements déformant de manière difficilement contrôlable les statistiques de naissances, je me suis abstenu de calculer les taux de fécondité légitime et illégitime. Cependant la longueur des séries temporelles obtenues pour la fécondité générale, la mortalité, et les migrations nettes me permettent d'étudier les déterminants de la transition démographique dans l'espace géographique. Pour cela, les techniques modernes de la co-intégration permettent de tester des causalités temporelles au sens statistique, et de contribuer au débat soulevé par Carlsson (1969) consistant à se demander si les changements de fécondité correspondent à une adaptation à un environnement changeant (la montée de l'instruction, l'urbanisation, la mobilité géographique, la progression de l'espérance de vie sont les variables disponibles en séries temporelles), ou bien si ces changements ont leur propre dynamique de diffusion d'une innovation dans l'espace et le temps.

En chemin, la place de la France dans la transition démographique européenne est révisée, l'existence d'un *baby-boom* au XIX<sup>e</sup> siècle est mise en question et imputée au sous-enregistrement des naissances, l'importance de la guerre de 1870-71 est réévaluée. Le résultat global est que le XIX<sup>e</sup> siècle réunit en

France un éventail large de trajectoires possibles. L'hétérogénéité des démographies départementales diminue en allant vers des régions de moins en moins rurales, jusqu'à disparaître pour des régions comme la Normandie, la vallée de la Garonne ou la Champagne. Ces dernières régions sont entrées, depuis le début du siècle, dans un régime de relative basse fécondité et basse mortalité, avec pour conséquence les populations les plus âgées (figure 8.14 page 130). La qualité de l'enregistrement est toujours très bonne dans ces régions, alors qu'elle s'améliore rapidement en Bretagne, par exemple, en partant d'une situation très mauvaise en début de siècle. Lorsque la sensibilité aux modifications de l'environnement cesse, il reste une dynamique spatiale et temporelle régie par la diffusion des nouveaux comportements moins féconds, et qui entraîne, au début du XX<sup>e</sup> siècle, une convergence des comportements selon les départements, un essor des migrations, et une homogénéisation de l'espace français.

*Présentation des données :*

- Vous pouvez télécharger les tables de mortalité en âge quinquennal par département et pour la France entière de 1806 à 1906, les tableaux de migration nette par âge quinquennal. Dans Bonneuil (1997), ces tableaux n'ont pas été publiés faute de place. Pour le tableau des espérances de vie par département et pour la France entière, le tableau téléchargeable est mis à jour par rapport à celui du livre.
- Les indices de fécondité de Coale figurent en annexe de Bonneuil (1997).

*Références :*

- « Reconstruire la population féminine de l'Hérault entre 1856 et 1906 », *Population* 3, 1998, 517-534.
- Bonneuil, Noël, *Transformation of the French demographic landscape*, 1997, Oxford : Clarendon Press.

**Annexe : Transformation des quotients de mortalité en quotients de mortalité perspectifs** Dans Bonneuil (1997, pages 80-81, je rappelle pour l'avoir utilisée la technique classique de passage des quotients de mortalité aux quotients perspectifs. Page 81, une coquille s'est glissée : il faut remplacer  $\bar{q}_{x-1}$  par  $\bar{q}_{x+1}$ . Il faut donc lire :

“for  $x = 1$  to 88 years old,

$${}_1q_x = 1 - \frac{(1 - {}_1\bar{q}_x)(2 - {}_1\bar{q}_{x+1})}{(2 - {}_1\bar{q}_x)}$$

”

Les tableaux téléchargeables ont été obtenus en remplaçant cette approximation *ad hoc* par un lissage bi-dimensionnel, qui ne figure donc pas dans Bonneuil (1997).

J'ai donc refait tous les calculs avec ce nouveau programme de passage en perspectif. Celui-ci consiste à construire une surface donc un axe est l'âge  $x$ , un deuxième le temps d'une année divisée en 10 intervalles égaux  $[t_i, t_{i+1}]$ ,  $i = 0, \dots, 9$ , et le troisième la fonction de survie  $l(x)$  à chaque âge exact  $x$ . Le lissage bi-dimensionnel permet de calculer la fonction de survie  $l(x)$  en tout point de la surface continue passant par les  $l(j)$ , où  $j$  est entier,  $j = 0, 1, \dots, 100$ . Il suffit ensuite de regrouper les  $l(x)$  suivant le découpage adéquat, et d'en déduire les quotients perspectifs. A mon avis, cette technique est plus précise que l'approximation classique reposant sur des coefficients *a priori*. Voici le programme correspondant en ratfor, faisant appel à des programmes IMSL.

```

subroutine perspn(qf)
#quotient annuel qf en entrée→
#passage en perspectif qf →
#redécoupage (facultatif) en quinquennal qf en
#sortie (de naissance a 0 – 4 rangé dans qf(1)
#de x, x + 4 à x + 5, x + 9 rangé dans qf(x))
integer age,i,j
real kn,qqq0,qqq,qqs0,qqs
dimension qf(0:95),qm(0:95)
integer kxord,kyord,ldf,nxdata,nxknot,nydata,nyknot
parameter (kxord=3, kyord=5, nxdata=11, nydata=95, ldf=nxdata,
nxknot=nxdata+kxord, nyknot=nydata+kyord)
integer nout,nxcoef,nycoef
real bs2vl, bscoef(nxdata,nydata),fdata(ldf,nydata),
float,sxy,x,xdata(nxdata),xknot(nxknot),y,ydata(nydata),yknot(nyknot)
intrinsic float
external bs2in, bs2vl, bsnak
do i=0,10
xdata(i+1)=float(i)/10. #1er axe : découpage d'une année #en 11 in-
tervalles égaux
do age=0,94
ydata(age+1)=age #2eme axe : l'âge
surv=1.
do age=0,94
{
do i=0,10
fdata(i+1,age+1)=surv #calcul de la fonction de survie
surv=surv*(1-qf(age))
}
call bsnak(nxdata,xdata,kxord,xknot)
call bsnak(nydata,ydata,kyord,yknot)
call bs2in(nxdata,xdata,nydata,ydata,fdata,ldf,kxord,kyord,xknot,yknot,bscoef)
nxcoef=nxdata
nycoef=nydata

```

```

surv1=1
do age=1,94
{
surv=0
do i=0,10
{
x=float(i)/10.
y=age-x
sxy=bs2vl(x,y,kxord,kyord,xknot,yknot,nxcoef,nycoef,bscoef)
surv=surv+sxy
# print *, "age,i,x,y,sxy,surv",age,i,x,y,sxy,surv
}
surv=surv/11.
if (age==1)
kn=1.-surv/surv1
else
qm(age-2)=1.-surv/surv1 #quotient annuel perspectif
# if (age==1)
# print *, "nais",surv,surv1,1.-surv/surv1,qf(0),kn
# else
# print *,age,surv,surv1,1.-surv/surv1,qf(age-1),qm(age-2)
surv1=surv
}
#passage en quinquennal
# naissances
qqq0 = 1.
do i=0,3
{
qqq = 1.
do j=0,i
qqq=qqq*(1-qf(j))
qqq0 = qq0 + qqq
}
kn=1.-(1-kn)*qqq0/5.
#autres âges
do age=0,80,5
{
qqq0 = 1.
qq0 = 1.
do i=0,3
{
qqq = 1.
qq0 = 1.
do j=0,i
{

```

```

qqq=qqq*(1-qf(age+j))
qqq0=qqq*(1-qf(age+5+j))
}
qqq0 = qqq0 + qqq
qqq0 = qqq0 + qqq
}
do i=0,4
qqq0 = qqq0*(1-qf(age+i))
qf(age)=1.-qqq0/qqq0
}
qf(1)=kn # commodit  d' criture
return
end

```