Life tables for France as a whole and by *département*, 1806-1906. Net migration tables by *département*, 1806-1906.

Noël BONNEUIL
Institut national d'études démographiques, 133 boulevard Davout, 75980 Paris Cedex 20, France, and Ecole des hautes études en sciences sociales
April 2005

In *Transformation of the French demographic landscape* (1997, Oxford: Clarendon Press) I challenged the reliability of the recapitulative mortality data in addition to that of censuses and birth statistics. I adduced the death by age statistics not used by Etienne van de Walle (*The Female Population of France in the Nineteenth Century*, 1974, Princeton: Princeton University Press).

The first task (Chapter 3) was to correct the data of the Statistique Générale de la France (SGF) for book-keeping accuracy, by cross-checking in the original documents with the sub-totals given for lines (*département*) and columns (age), and the tables published by marital status, or for "both sexes". Although this initial stage did not involve any reconstruction of the censuses, let alone judgement of quality, it did give me sound data, accurate on a book-keeping basis, to start from. All the information available in SGF is now incorporated in the reconstruction in a way that is demographically consistent. This means that all the demographic equations must be verified, but only those equations. The method comprised the following stages:

1. In each *département* i, the set of individuals of a given age x at date t is subjected to two concurrent forces between date t and date $t+dt$: death $\mu(x, t)$ and emigration $\nu(x, t)$ (equation 5.1, page 57). With a technique invented by Greville (1948), and used by Chiang (1968) and Brouard (1986) among others, I integrated the equation on the basis of the weakest calculation hypothesis (see discussion in Schoen 1988). I showed that it is also possible to incorporate immigrants. For a period of time h, probability of dying may be expressed as ${}_hq^i_x$ between ages x and $x+h$ as a function of the total number of deaths $D_i([x, x+h])$ between t and $t+h$, and the population figures $p_i(x, t)$ and $p_i(x+h, t+h)$ at the censuses held at t and $t+h$. These figures are taken from movements and censuses of variable quality (e.g., the 1861 and 1866 censuses are good, but 1872 is poor). This is equation 5.5 (Bonneuil, 1997, 62). However, $D_i([x, x+h])$ is unknown, and $p_i(x, t)$ and $p_i(x+h, t+h)$ are taken from censuses of variable quality.

2. Death totals for a set age group, for example [10, 14] years, were published each year beginning in 1856. The totals consequently involve a number of cohorts. In Bonneuil (1997, 47-52) I undertook to isolate deaths for each cohort within the age-group total. To that end, I used a national age and month distribution published fairly regularly along with the *département* death figures by age. Then applying two-dimensional smoothing to deaths by age, in such a way as to maintain the total number of deaths per age group, each death was allotted to a specific cohort. This numerical technique avoids the use of ready-made formulae, such as calculating cohort coefficients from mortality rates of individuals of various ages. This latter approximation may work for closed communities with little irregularity in generation numbers and mortality, but for 19th-century French *départements*, the aim is precisely to describe open communities with fluctuations in birth and death numbers.

3. Examination of the $p_i(x, t)$ numbers (Van de Walle 1974, 36, fig. 2.6, quoted in Bonneuil 1997, 44-45) shows that they may be under-estimated or distorted by digit preference or age heaping, or even miscounting (especially in the 1872 census). Pre-processing avoids these difficulties. It provides $p_i(x, t)$ and $p_i(x+h, t+h)$ estimates that are incorporated in equation 5.5 in Bonneuil (1997, 62). In this way the probabilities of dying ${}_hq^i_x$ are obtained. This does not prejudge the final $p_i(x, t)$ estimates.

4. To allow for age heaping, a number of techniques are proposed, as for example in UN Manual X.[1] It is generally hard to distinguish between an age heaping effect and the passing of a more numerous generation. In our case, having a series of censuses every five years from 1851 to 1906 makes it possible to resolve the ambiguity: I suggested first identifying the cohorts in each census[2] and then observing cohort numbers. These numbers regularly fluctuate with peaks at certain age digits and troughs at others. Van de Walle noted this phenomenon and excluded the effect of migration. I fully accepted his point and suggested a least-squares smoothing procedure. Figures 4.1, 4.2, 4.3, A1, A2 and A3 (Bonneuil 1997, 41-43, 174-76) show that this smoothing carefully corrects the irregularities in the examples of the Creuse and Finistère *départements*. Figure 1 represents the Charente *département* for cohorts 1826-30 to 1836-40.

[1] Manual X: Indirect techniques for demographic estimation. New York: United Nations, 1983.

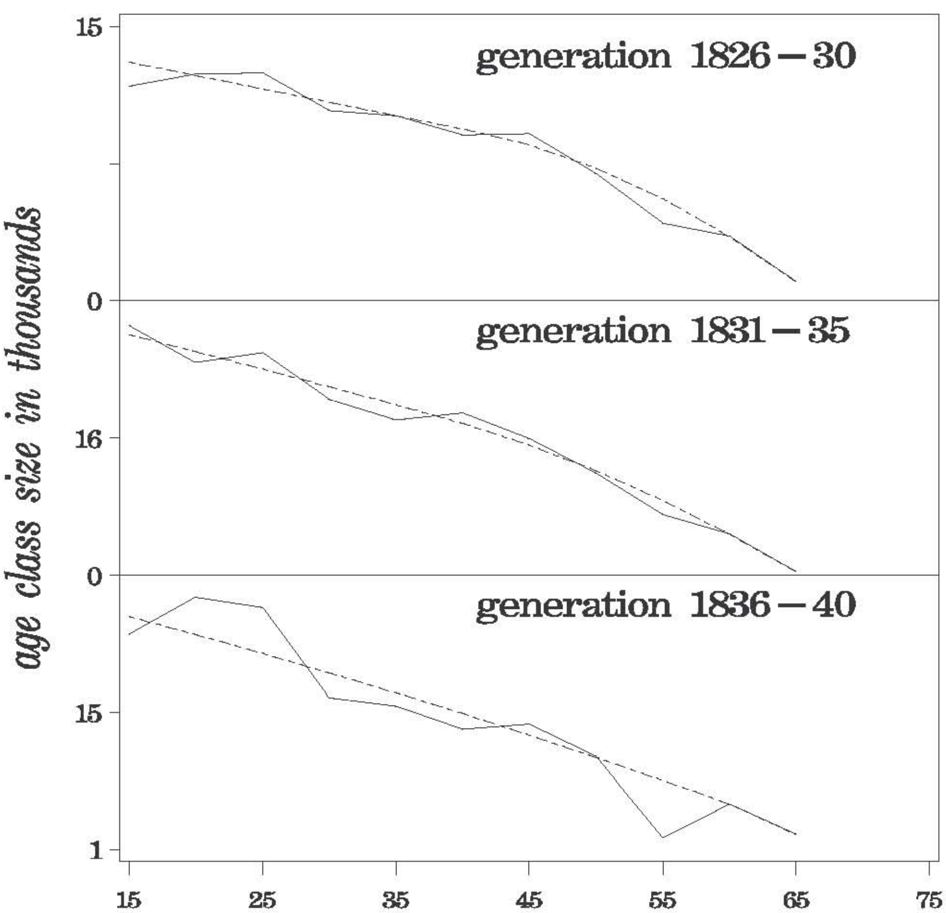
[2] The initial adjustment is required because the intercensal period is not exactly five years, but may be four and a half (December 1881 to May 1886) or even six years, as from 1866 to April-May 1872.

5. The probabilities of dying ${}_hq^i{}_x$ between t and $t+h$ are calculated for each age (from birth to 0-4 years, from 0-4 to 5-9 years, etc.), each quinquennial period, and each *département* independently, except for $p_i(x, t)$, which is involved, even if corrected meanwhile, in estimating ${}_hq^i{}_x$ ([t, $t+h$]) and ${}_hq^i{}_x$ ([t-h, t]). An initial plausibility test is to construct life tables estimated in this manner for each *département* and period, and compare them with model tables, such as Ledermann's. The Charente is given as an example in Figure 2, and Finistère, Creuse, Gironde and Seine in Bonneuil (1997), and Hérault in Bonneuil (1998). The shift from a table of probabilities of dying to a table of probabilities of dying within a calendar period is shown in the appendix. The fit with a single-parameter Ledermann table according to a chosen distance (see Bonneuil 1997, 63, 64 for distance selection) reveals both an improvement in data quality from the 1856 census onwards and a fairly systematic discrepancy at birth and 0-4 years. Although adjusting to a single-parameter model table is less flexible, it does provide a simple, if imperfect, way of assessing under-recording, which is most common for the younger ages. The Ledermann probability may well be lower than that directly estimated, as happens for some years, especially for the Haute-Vienne. In these very few cases, one may suppose either double counting or infant mortality that is actually higher than the closest Ledermann figure, with nil under-recording of births.

The directly estimated tables incorporated biased data ($p_i(x, t)$ and Di ([x, $x+h$]) numbers). One way of correcting them and achieving demographic consistency is to replace these tables by the Ledermann model tables that are closest in terms of minimising the chosen distance. Once this is done, the $p_i(x, t)$ values may be estimated again from the $p_i(x+h, t+h)$ values and ${}_hq^{i,}{}_x{}^L$([t, $t+h$]) Ledermann probabilities that have replaced the directly estimated probabilities ${}_hq^i{}_x$ ([t, $t+h$]). The demographic equations are then satisfied with no need for ad hoc restrictions, as close as possible to the data, without over-estimating their quality.

6. Equation 5.1 in Bonneuil (1997, 51), expressing variation in the $p_i(x, t)$ number, is used to calculate the probability of dying and the net migration rate (equation 5.7, p. 62) by age, *département* and quinquennial period independently (except for the $p_i(x, t)$ numbers which are involved in two equations, as mentioned above). The trajectory by time and age of the various net migration rates shows great consistency within each *département*, which is an ex post validation of the reconstruction, as is the close fit between the directly estimated life tables and Ledermann's model tables.

Figure 1: Least-square smoothing by cohort to correct age heaping. Charente, generations 1826-30 to 1836-40.



age class size in thousands
15
0
16
0
15
1
generation 1826 — 30
generation 1831 — 35
generation 1836 — 40
15
25
35
45
55
65
75
Age

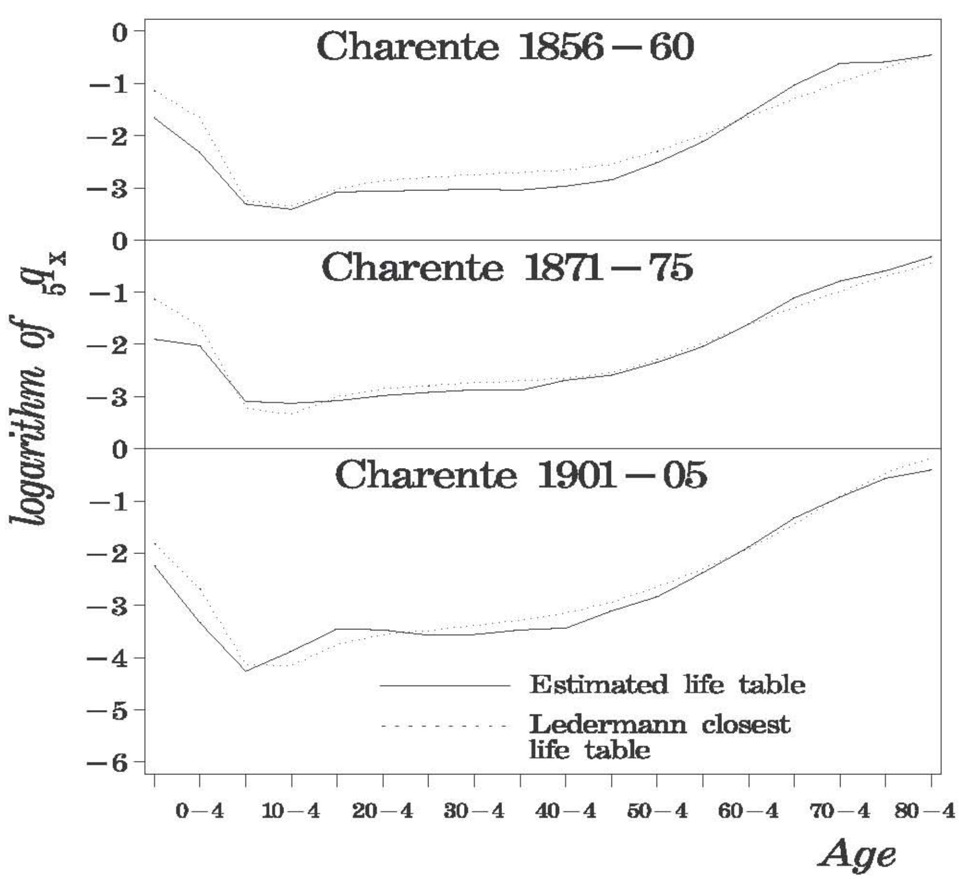

Charente 1856−60
Charente 1871−75
Charente 1901−05
logarithm of $_5q_x$
0
−1
−2
−3
−4
−5
−6
Estimated life table
Ledermann closest life table
0−4
10−4
20−4
30−4
40−4
50−4
60−4
70−4
80−4
Age

Figure 2: Example of adjustment of probabilities of dying reconstructed with Ledermann probabilities. Charente.

7. The results depend on the techniques used, such as preliminary smoothing to correct age heaping. I give a sensitivity analysis on pages 67-72.

8. From 1806 to 1856, death by age statistics were not published. The reconstruction from 1856 to 1906 provided not only net migration rates by age but also a criterion for correcting the estimated model tables. I used these two results to reconstruct the 1806-1856 period. In this way I deliberately bias the model tables used by applying to 1806-1856 the bias observed in 1856-1906, and extrapolating backwards net migration rates by age and *département*. By considering that the age pyramid in 1856, obtained by forward reconstruction from 1806 to 1856, must be identical to that obtained by step-wise backward reconstruction from 1906 to 1856, it is possible to propose corrections of both the 1806 census and births from 1806 to 1855 (the other censuses in this period are not usable). It so happens that the trend line in estimated under-recording from 1806 to 1856 is the same for each *département* as that from 1856 to 1906, although the two methods used are quite different. The 1806 census correction raises the estimated female population between 1806 and 1851 (15.2 million in 1806, compared with van de Walle's estimate of 15.0, and 18.5 million in 1851, compared with van de Walle's 18.3 for France as a whole).
9. Since under-recording distorts birth statistics in a manner that is hard to check, I have deliberately not calculated legitimate and illegitimate fertility rates. However the length of the time series obtained for general fertility, mortality and net migration has meant that I have been able to study the determining factors in demographic transition in a given geographical area. Modern co-integration techniques are used to test temporal causalities in the statistical sense, and to contribute to the debate begun by Carlsson (1969) on whether changes in fertility correspond to an adaptation to a changing environment (improved education, urbanisation, geographical mobility, increased life expectancy are variables available in time series) or whether these changes have their own dynamic as the dissemination of an innovation in time and space.

As a result, the place of France in the European demographic transition is revised, the existence of a 19th-century baby boom is challenged and imputed to under-recording of births, and the impact of the 1870-71 Franco-Prussian War is reassessed. The overall result is that 19th-century France displayed a wide range of possible trajectories. The variations in *département* demographics decline in the less rural regions, disappearing altogether in places such as Normandy, the valley of the Garonne and Champagne. These latter regions entered a system of relatively low fertility and low mortality at the start of the century, giving them the oldest population groups (fig. 8.14, p. 130). Record quality was very high in these regions throughout the century, whereas Brittany, for example, improved rapidly from a very poor position at the start. When sensitivity to modifications in the environment ceased, there remained a dynamic in space and time that was driven by the dissemination of new behaviours of lower fertility, leading by the turn of the 20th century to a convergence in behaviour among *départements*, a boom in migration, and a degree of uniformity across France.

Presentation of the data:

- Available online are life-tables by five-year age group for each *département* and France as a whole from 1806 to 1906, net migration tables by five-year age group. They were not published in Bonneuil (1997) for reasons of space. In the case of the life expectancy table for each *département* and France as a whole, the online table has been updated since the book was published.
- The Coale fertility indices are given in the appendix to Bonneuil (1997).

References:

- 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.

Appendix: Converting probabilities of dying to probabilities of dying within a calendar period. In Bonneuil (1997, 80-81), I recapitulate from my own use the standard technique for going from probabilities of dying to probabilities of dying within a calendar period. On page 81 there is a misprint: $\bar{q}_{x-1}$ should be replaced by $\bar{q}_{x+1}$. It should read:

"for x = 1 to 88 years old,

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

The online tables were obtained by replacing this *ad hoc* approximation by a two-dimensional smoothing that is not in Bonneuil (1997).

I have repeated the calculations using the new programme for conversion to probabilities of dying within a calendar period.

This consists of constructing a surface of which one axis is age x, the second years divided into 10 equal intervals $[t_i, t_{i+1}]$, $i = 0, \ldots, 9$, and the third survival function $l(x)$ at each exact age x. Two-dimensional smoothing is used to calculate the survival function $l(x)$ at all points on the continuous surface passing through points

$l(j)$ where j is an integer from 0 to 100. Then the $l(x)$ points need only to be collected into convenient groups and the deduced probabilities of dying within a calendar period. In my view, this technique is more accurate than the standard approximation using a priori coefficients. Below is the corresponding ratfor (fortran77) programme, using IMSL programmes.