PERIOD FERTILITY MEASURES
Reflective commentaries

Demographic analysis was for a long time restricted to the measures which could be derived directly from annual statistics. Even the retrospective information contained in population censuses (children already born, proportions single) was neglected. The possibility of constructing cohort measures, and the simplicity of such indicators, did not emerge until the post-war years. The primacy given to these summaries of cohort lifetimes then resulted in a preference for those period measures which fit in best with the longitudinal approach. Rates calculated like age-specific fertility rates and summed cross-sectionally were preferred, since it was shown that they reflected changes over time in both the quantum and the tempo of cohort fertility**. Jean-Louis Rallu and Laurent Toulemon (JLR and LT) have, on the contrary, taken the stand that period measures should be as free as possible of the influence of the cohorts' past, and have investigated alternative period fertility constructs. We solicited the reactions of a number of colleagues to various points they raise. Their remarks are published here, followed by the authors' reply.*

SYNTHETIC MEASURES BASED ON RATES OR ON PROBABILITIES

Gérard CALOT***

The study of change through time in phenomena which, at the individual level, also develop through time, lies at the heart of demographic analysis: fertility, nuptiality, mortality change, etc. It raises many thorny problems. We propose to discuss a few of them.

The time markers Each event can be located in time by two dates: one referring to the individual concerned (or unit, e.g. the couple), the other to the actual instant of occurrence. *Longitudinal*

* See L. Henry, «Réflexions sur la conjoncture», *Population*, 6, 1991.
** See G. Calot, «Relationships between cohort and period demographic indicators. The translation problem revisited», *Population: An English Selection*, 5, 1993.
*** Former Director of INED.

Population: An English Selection, 6, 1994, 95-130

analysis focuses on the former and studies events occurring in *successive cohorts,* the time marker being the calendar year when the cohort was formed. *Transversal* analysis deals with the latter, and studies events which have taken place in *successive periods*, the time marker being the calendar year of occurrence.

The cohort approach is *a priori* more attractive, because more natural. It follows time exactly as it unfolds and defines groups – cohorts – which closely resemble individuals, since the characteristics of the group are simple aggregates of those of its individual members (averages or proportions). A cohort's completed fertility is the *average* number of children borne by its members, the parity progression ratio p_2 is the *proportion,* among the members who have had at least 2 children, of those who have at least 3.

But, fundamentally, the question is of the *pertinence*[1] of the successive groups thus defined: to what extent does a cohort represent a concrete reality? Is there not greater homogeneity among the set of events occurring within a same period than within a same cohort?

The understanding of changes through time is probably more immediate across cohorts than across periods, but how does the causal relationship work? Is the cohort the cause and the period the effect or is it not more – or more often – the other way around? Conceptually, the longitudinal perspective seems more natural, but the world we live in is essentially periodwise. Also, if there is an overall longitudinal consistency at the individual level, does that imply that it necessarily exists at the cohort level? Finally, what is more important from the point of view of the population studied: period developments or cohort developments?

The variables The person for whom an event is recorded bears a number of characteristics: first, the difference between the two time markers, that is, age (or duration in status) at time of occurrence; second, any other descriptive or potentially explanatory variable, that is, corresponding to different frequencies of occurrence. Here, the range is considerable: at one end, the variables which are *unchanging* for the individual (sex, place of birth), *barely changing* or changing *independently* of the event observed (socio-occupational category, educational status, region of residence), at the other, the *endogenous* variables which *automatically change* when the event studied occurs (marital status in the case of nuptiality, number of children already born for fertility by birth order...). The former give rise to differential analyses, while the latter are part and parcel of the analysis of the process leading up to the event considered.

The more variables we analyse, the greater the volume of data to be processed and also – when working on samples – the uncertainty, owing to sampling errors.

---

(1) On this important issue, see Máire Ní Bhrolcháin, 1992.– «Period paramount? A critique of the cohort approach to fertility», *Population and Development Review*, 4, 599-629.

Synthesis of period observations The tools of statistical observation used for cohort analysis are naturally the same as for period analysis. Cohort rates or probabilities aggregate naturally into *synthetic measures* of cohort quantum: *summing* age-specific first marriage rates (incidence rates): $T=\sum_i t_i$ or combining age-specific first marriage probabilities, by *multiplying their complements*: $1-Q=\prod_i (1-q_i)$, result in the same proportion ever-married by age 50.

Similarly, when period rates are applied to the population *structures* specific to the period, they lead to the same total number of events occurring during the period.

But we prefer to synthesize period quantum *independently* of these specific population structures – which are, in a way, *contingent* – to obtain a result expressed, if possible, in cohort terms, closer to the individual. To do so, we concoct the *fictitious cohort*: we determine what the longitudinal measure would be for a cohort of women whose fertility performance at each different stage of their life course would be that of the period considered.

Naturally, the result obtained will depend on the *number and nature of the variables selected.* When we add another one *to a given set of variables*, we take into account the changes affecting the population structure specific to this new variable, for each combination of the other variables. In other words, we move the dividing line which separates variation in the total number of events observed into structural effects and quantum effects. For instance, when first marriage in a specified year is summarized by cumulating the age-specific first marriage *rates* (incidence rates), the effects eliminated are those of change in the age structure of the *total* population. When we combine the age-specific first-marriage *probabilities*, the effects eliminated are those of change in the age structure of the *single* population. In the first case, we take into account only the age structure of the total population; in the second, we *also* consider the cumulated effect of nuptiality in years prior to the period observed.

We can then wonder to what extent it is licit, and useful, to keep on adding extra variables, given that this increases the cost of the calculations, while their transparency, and the possibility of performing them for a wide range of countries or periods, diminishes. Furthermore, when can one set of variables be said to be preferable to another?

In the above example – first marriage – it seems better, intuitively, to derive the synthetic measure from the probabilities: past nuptiality being necessarily an explanatory factor of present nuptiality (the transition from never-married to married can only take place once), its effect should be eliminated. But does the fictitious cohort principle actually eliminate it correctly?

Age-specific first marriage

Let us examine in detail the case of age-specific first marriage. We denote: $P(i–1,n)$ the total population of age $i–1$ on January 1st of year n; $c(i–1,n)$ the proportion single in this population; $t(i,n)$ the first marriage rate at age i in year n (incidence rate); $q(i,n)$ the first marriage probability at age i in year n; $M(i,n)$ the number of first marriages at age i in year n. We assume the absence of disturbing events (mortality, migration).

The total number of first marriages in year n before age 50 is:

$$M(n) = \sum_{i}^{49} M(i, n)$$

$$= \sum_{i}^{49} P(i - 1, n) \cdot t(i, n)$$

$$= \sum_{i}^{49} P(i - 1, n) \cdot c(i - 1, n) \cdot q(i, n)$$

What *would have been* the number of first marriages had the population structures on January 1st not been those specific to the year considered, but had coincided with simple structures, defined a priori, and time (and space) *invariant*?

• In the case of the populations $P(i–1, n)$, it seems reasonable to assume that the number of marriages, at each age, would have been *proportional* to what it was had the population of that age been different. It is true that there is 'competition' on the 'marriage market' and that substitutions can be imagined with other age groups, or other countries, when the male and female age structures are unbalanced. We shall put these considerations aside and assume that, had the populations $P(i–1,n)$ been those of a *uniform* pyramid, composed of the same number of persons of each age (the number, being then meaningless, is put at 1 by convention), the number of first marriages before age 50 would, 'all other things being equal', have been:

$$M_1(n) = \sum_{i}^{49} t(i, n)$$

M_1 is equivalent to the TFR index in the study by JLR and LT. It aggregates the age-specific incidence rates *by summation*.

• It is much more problematic to determine what the number of first marriages at age i would have been had the 'stock' of single persons on January 1st been, in proportion, different from what it was, $c(i–1, n)$. Indeed, the observed probability $q(i,n)$ is not independent of $c(i–1, n)$: 'all other things being equal' for the propensity to marry, it is even reasonable to suppose that the greater the stock, the higher the probability expressing the *same* propensity to marry as $q(i,n)$ associated with $c(i-1,n)$. But how much higher?

At this stage of our reflection, we can either:

— refuse to forward an estimate of what $q(i,n)$ would have been, and stay with M_1, which adjusts the number of first marriages for the only factor of variation that can reasonably, and more or less satisfactorily, be adjusted: the size of the *total* (not the exposed) population of each age;

— go ahead despite the above objection, acknowledging that we don't really know the significance of a given probability variation when the value of $c(i-1,n)$ also varies, but accepting this as of minor importance when marriage time patterns do not change too much. The number of first marriages before age 50 then becomes, supposing that replacing $c(i-1,n)$ by $c_0(i-1)$ would not have altered the probability $q(i,n)$:

$$M_2(n) = \sum_{i}^{49} c_0(i-1) \cdot q(i, n)$$

$c_0(i-1)$ being the given structure at age $i-1$ (standard population method);

— go ahead despite the above objection, and consider that 'the nuptiality conditions in year n' is an expression that makes sense. In this case, the 'singles stock' on January 1st of year n is the proportion single that would have been observed if, *in the years preceding n, the age-specific first marriage probabilities had been those in year n*; then we apply to this stock the *probability $q(i,n)$ actually observed:*

$$M_3(n) = \sum_{i}^{49} c^*(i-1, n) \cdot q(i, n)$$

where

$$c^*(i-1, n) = \prod_{j}^{i-1} [1 - q(j, n)]$$

Then $M_3(n)$ can be written:

$$M_3(n) = 1 - \prod_{i}^{49} [1 - q(i, n)]$$

This calculation corresponds to the *fictitious cohort principle* and the PATFR index in JLR and LT (with only one possible order in the case of *first* marriages).

We note that the three indicators $M_1(n)$, $M_2(n)$ and $M_3(n)$ which, with the exception of the second, are expressed in terms of 'proportion ever-married at age 50', are of the form:

$$\sum_{i}^{49} \alpha(i) \cdot q(i, n)$$

where

$$\alpha(i) = \begin{cases} c(i-1, n) & \text{for } M_1 \\ c_0(i-1) & \text{for } M_2 \\ c^*(i-1, n) & \text{for } M_3 \end{cases}$$

How the indicators react to sharp change in the probabilities

We say that, in years m preceding n, nuptiality conditions have *not changed*: the 'stocks' $c(i–1,m)$ are *constant* at each age, and so are the rates $t(i,m)$ and the probabilities $q(i,m)$. Thus, $M_1(m)$, $M_2(m)$ and $M_3(m)$ are constant, which is satisfactory.

Suppose that, in year n, the probabilities at each age i suddenly exceed their previous values by (say) 10%. How will this change be reflected in the indices?

In M_1 and M_2, it is *faithfully* reflected: both mark a 10% increase. But M_3 rises by less than 10%, since, in the sum:

$$M_3(n) = \sum_{i}^{49} c^*(i-1, n) \cdot q(i, n)$$

$q(i,n)$ does increase by 10%, but $c^*(i–1,n)$ *decreases*:

$$c^*(i-1, n) = \prod_{j}^{i-1} [1 - q(j, n)]$$

Consequently, M_3 does not reach +10%, and in fact, *by a long way* when the phenomenon under study is, as first marriage was not so long ago, not far from universal (see Appendix I).

In this example, the *right* answer is obviously 10%: this is the *specific* variation in the number of marriages produced by variation in frequency, without any change in the population of each age. In this respect, M_3 is a *poor* indicator of the specific variation in the number of marriages. But what is the exact significance of M_3 and of its relative variation?

$M_3(n)$ is the proportion ever-married at age 50 that would *ultimately* be observed if the age-specific probabilities in year n were repeated *persistently*. Thus, the relative variation of M_3 is the relative variation in the proportion ever married at age 50 in this radically 'conservative' perspective, of which we know neither the *realism* (is it possible?) nor the *plausibility* (is it probable?). In fact, there is no reason why such 'ad vitam æternam' should be contemplated, or referred to. But, even if we do not know exactly what a 10% change in probabilities signifies, at least the effect on the number of marriages is correctly measured by M_1 and M_2.

Generally speaking, it seems essential that the relative variation in the synthetic index be contained between the extreme relative variations of the age-specific rates or probabilities (as is the case with a weighted average): the summary indices M_1 and M_2 satisfy this condition, but not M_3 (see Appendix II).

The shortcomings of M_3 come into full light when we take the example of mortality, a universal phenomenon (or fatal event). Suppose that the probabilities of dying at each age are 10% lower in year n_2 than in n_1. M_2, being based on a population pyramid which is realistic albeit conventional, will faithfully reflect this reduction, which is naturally echoed by the absolute number of deaths in identically sized age groups (M_1). M_3, on the contrary, being based on the fictitious cohort principle, will be totally insensitive to such change in the probabilities: its value will invariably be 1, expressing the fact that, even with a 10% drop in the probabilities of dying, man remains mortal!

In conclusion, a period synthesis of probabilities based on the fictitious cohort principle is in no way an improvement on a period synthesis of incidence rates (when a standard population can be easily defined, which is the case for fertility or first marriage with the uniform age structure). There is, in fact, every reason to believe the opposite.

The exact significance of the period syntheses

What matters is to understand the exact significance of the period indicators. In particular, it is important to note that the measure derived from the first marriage *rates*, which is regrettably expressed in terms of proportion ever-married at age 50, should not be interpreted in these terms, but as a measure of the *absolute number of first marriages*, adjusted for the effect of changes in the age structure of the *total* population.

In 1990, for instance, the synthetic measures of female first marriage in France amounted to:

— derived from the rates: 0.56

— derived from the probabilities: 0.71

whereas both measures were in the region of 0.92 in 1970. The exact significance of these results is as follows:

• in the space of twenty years, the specific effect, on the number of female first marriages before age 50, of the change in the age-specific first marriage rates, given that the age structure of the *total* female population has not varied, is a decrease of:

$$1 - \frac{0.56}{0.92} = 39\%$$

This means that, *on average*, the female first marriage *rates* for the same age fell by 39% between 1970 and 1990 (see Appendix II): taking into account their size, the cohorts of marrying age had produced 39% fewer brides in 1990 than in 1970. Or that the number of marriages in 1990, owing to nuptiality in 1990 but also in the preceding years, was

39% lower than in 1970, after adjustment for change in the size of the cohorts concerned in 1970 and 1990;

• were the age-specific female first marriage probabilities observed in 1990 maintained indefinitely, this would *ultimately* lead to a proportion of 71% of women ever-married at age 50.

This means that the first marriage probabilities at each age have, *on average*, increased in the space of twenty years by (cf. final equation in Appendix I):

$$\delta = \frac{\ln(1-0.71)}{\ln(1-0.92)} - 1$$

that is, here, have *diminished* by:

$$-\delta = 1 - \frac{\ln(1-0.71)}{\ln(1-0.92)}$$

$$= 51\%$$

In other words, the specific effect, on the number of female first marriages before age 50, of change in the age-specific first marriage probabilities, given that the age structure of the *single* female population has not varied, is a decrease of 51%: taking into account their numbers, the single women of these ages produced 51% fewer marriages in 1990 than in 1970.

The difference between 39% and 51% expresses the fact that on January 1st 1990, there were relatively more single women close to the modal marrying age (25) than on January 1st 1970, due to lower nuptiality before age 25 in the years preceding 1990 than in those preceding 1970. In this respect, 51% is perhaps a better measure than 39% of the nuptiality decline during these twenty years: while the transition of M_3 from 0.92 to 0.71 between 1970 and 1990 should not be interpreted as the transition from the 'first marriage conditions' corresponding to 92% of ever-married at age 50 to those corresponding to 71%, the equation we have used:

$$\delta = \frac{\ln[1 - M_3(n_2)]}{\ln[1 - M_3(n_1)]} - 1$$

based on variation in M_3 seems to yield a better measure of change in the phenomenon studied than the simple relative variation of M_1. However, we can also say that, in the comparison between 1970 and 1990, we are dealing with a *minimum*. Indeed, if, at each age, the first marriage probability in 1990 had come back to its level in 1970, whereas the 'stock' of eligible women of each age was relatively larger than in 1970, the relative variation of M_3 would have been nil, and yet that would have meant probably less favourable 'first marriage conditions' in 1990 than in 1970...

In fact, we prefer not to use the term 'first marriage conditions' in 1990 or 1970, because we cannot tell to what extent the assumption that female first marriage probabilities could persistently repeat the values observed in a given year is *realistic* or *plausible*(2). Furthermore, although the sets of probabilities in 1970 and 1990 were actually observed, each one reflects its own specific *past*. 'Freezing' the probabilities amounts to gradually changing the past, but keeping the same probabilities for the present, which is meaningless unless the past has no influence on the present. Now, two first marriage probabilities which are equal, but applied to substantially different 'stocks' of singles, by no means have the same significance(3). The 'all other things being equal' line of reasoning which is apparently held when one freezes the probabilities, is far from being pertinent and is, in fact, a fallacy, in times of rapid change.

*

* *

But does that mean that summing a set of age-specific rates should be considered a satisfactory way of summarizing period fertility?

Again, the method is based on the fictitious cohort principle, since the age-specific rates are those actually observed in a given year n, but each one in a different cohort. The result obtained is a summary measure of the cohort performance that would *ultimately* be observed if, through time, each age-specific rate *persistently* held its value in year n. But, as we shall see, it is possible to give another reading to this measure, which does not require imagining any such permanency.

Let us consider, for instance, how generations succeed one another: what does the number of births in a given year mean in the perspective of *generation replacement* [1]?

We know that, in year n, the total number of births is the product of the total period fertility rate (denoted I) and the average size of the female cohorts of childbearing age (G):

$$N(n) = I(n) \cdot G(n)$$

The average size of the female cohorts of childbearing age in year n is, to be precise, the weighted average of the numbers of women of each age, the weights being the fertility rates *in year n*. But when the size of the cohorts varies only slightly, or fairly regularly, the weights have very little influence on the value of the weighted average: they might just as well be the fertility rates observed another year, or any realistic fertility schedule. Consequently, the average size of the female cohorts of childbearing age $G(n)$ has an intrinsic significance, which is practically independent of the set of rates adopted. Thus, if we have a projection of the female populations of individual ages 15 to 50, we can project $G(n)$ from their simple weighted average.

(2) An assumption of this nature is *realistic* only if, in the mid term, the probabilities at each age have *not changed much*, and *plausible* when they have *barely changed at all*.

(3) When the 'stocks' are very different, identical probabilities do not express stability!

This average cohort size is that of the *generation of mothers*. The *generation of daughters* born in year n is equal to $\frac{100}{205}N(n)$, given a sex ratio of 105 baby boys per 100 baby girls. The ratio of the number of daughters – observed *at birth* – to mothers, who by definition are of the mean age at childbearing, about 28 years, is then:

$$r_0(n) = \frac{\frac{100}{205}N(n)}{G(n)} = \frac{100}{205}I(n)$$

Assuming that the female probability of survival from birth to this age is 0.985[4], the same ratio, measured at the time when the *daughters have become the population of reproductive age* is:

$$r(n) = \frac{98.5}{205}I(n)$$

$$= \frac{I(n)}{2.08} \neq \frac{I(n)}{2.1}$$

Thus, the period indicator $I(n)$ is identical, to a factor of 2.1, to the *ratio of the number of newborn girls to the average size of the cohorts who mothered them.* For *whatever reason*, be it due to changes in completed fertility or in fertility timing, *the cohort just born* – after adjustment for the age difference between mothers and daughters, i.e. taking mortality, but not migration, into account – *is, in terms of numbers, in a ratio of* $\frac{I(n)}{2.1}$ *with the average size of the cohorts having produced it*: *$I(n)$ is an indicator of the degree to which a generation born in year n replaces the generation having mothered it*, its base being not the conventional 100 in a given year, but 2.1, the point where replacement level is *reached exactly, international migration disregarded.*

In other words, in about 27 years (the modal age at childbearing), the size of the cohort contributing *most* to fertility will, compared to the *average* cohort today, be multiplied by $\frac{I(n)}{2.1}$ *(international migration disregarded). In France, for instance, the total fertility rate has fluctuated around 1.8 – almost 15% below 2.1 – for the last 15 years. As a result, the average size of the cohorts of childbearing age has started to decrease in 1991 (G =* 430,000) and should lose a further 10% between now and 2005 (cf. Figure 3 in reference [1]), in the absence of international migration.

Let us consider a final way of reading these results. We term *invariable fertility regime* one where, from one generation to the next, both *completed fertility* and *fertility timing* are unchanging. In such a regime, where D is the completed fertility value, the *rate of replacement of the childbearing cohorts*, i.e. the ratio of the number of females born in a

[4] This value corresponds to current mortality rates.

given year to the weighted average of the numbers of women of each childbearing age in the same year, is invariable and equal to $\frac{D}{2.05}$.

To say, in a context other than this invariable fertility regime, that the period fertility measure amounts to I in year n, therefore *means that replacement of the childbearing cohorts occurred, in year n, at the same rate as if an invariable fertility regime had prevailed, with completed fertility D equal to I.*

These reflections give a very concrete significance to the period fertility indicator (here denoted I, TFR in the article by JLR and LT): *it measures the degree of generation replacement.* As an instrument for assessing contemporary situations, it is consequently of considerable interest – which is not the case with the PATFR, PDTFR and PADTFR indices presented by JLR and LT. Furthermore, it is easily available for a great number of countries and periods, requiring nothing more than the age-specific fertility rates. It is even possible, supposing both mortality and international migration to be stable, to derive it with satisfactory accuracy from the total number of births [2]. Therefore, until there is evidence to the contrary, it is indeed, in the sense defined above, 1.8 children per woman that captures the fertility situation in France since 1975.

Gérard CALOT

REFERENCES

[1] Calot G.– «La relève des générations», *Population et Sociétés*, 265, février 1992.
[2] Calot G.– «Une notion intéressante : l'effectif moyen des générations soumises au risque», *Population*, 6, 1984 and 1, 1985.

APPENDIX I

Variation in M_3 and relative variation common to all age-specific probabilities

In continuous language, the indicator $M_3(t)$ at moment t is:

$$M_3(t) = 1 - e^{-\int^{50} q(x,\, t)dx}$$

where $q(x,t)$ is the first marriage probability at age x during the time interval $(t,t+dt)$. If we replace $q(x,t)$ by $(1 + \delta)q(x,t)$ (that is, all the probabilities increase by relative value δ), M_3 becomes:

$$M_3 + \Delta M_3 = 1 - e^{-\int^{50} (1 + \delta)\, q(x,\, t)dx}$$

Whence:

$$\Delta M_3 = e^{-\int^{50} q(x,\,t)dx} - e^{-(1+\delta)\int^{50} q(x,\,t)dx}$$

$$= (1 - M_3)\,[1 - (1 - M_3)^{\delta}]$$

When M_3 is of the order of 0.9 (10% single at age 50) and δ is of the order of 0.1 (relative increase of 10% for all probabilities), the relative increase in M_3 is:

$$\frac{\Delta M_3}{M_3} = \frac{0.1}{0.9}\,[1 - 0.1^{\,0.1}]$$

$$= 2.3\%$$

For a phenomenon like first marriage which is not very far from universal, variation in M_3, of the same sign as the variation δ common to all probabilities, may be far lower than δ in absolute value. It is only for rare phenomena (M_3 not nil but close to zero) that the relative variation of M_3 neighbours δ. The preceding equation can also be written:

$$(1 - M_3)^{\delta} = 1 - \frac{\Delta M_3}{1 - M_3}$$

that is:

$$\delta = \frac{\ln\left[\dfrac{1 - (M_3 + \Delta M_3)}{1 - M_3}\right]}{\ln(1 - M_3)}$$

APPENDIX II

Relative variation in the period indicator and in the rates or probabilities

The period first marriage indicator derived from the age-specific incidence rates is:

$$M_1(n) = \sum_i^{49} t(i,\,n) = \sum_i^{49} c(i-1,\,n)\cdot q(i,\,n)$$

whereas that derived from the probabilities by the standard population method is:

$$M_2(n) = \sum_i^{49} c_0(i-1)\cdot q(i,\,n)$$

The relative variation in these two indicators between the years n_1 and n_2 (which are not necessarily consecutive) is:

$$\frac{M_1(n_2) - M_1(n_1)}{M_1(n_1)} = \frac{\sum_i^{49} t(i, n_2) - \sum_i^{49} t(i, n_1)}{\sum_i^{49} t(i, n_1)}$$

$$= \frac{\sum_i^{49} t(i, n_1) \left[\frac{t(i, n_2) - t(i, n_1)}{t(i, n_1)}\right]}{\sum_i^{49} t(i, n_1)}$$

and:

$$\frac{M_2(n_2) - M_2(n_1)}{M_2(n_1)} = \frac{\sum_i^{49} c_0(i-1)\, q(i, n_1) \left[\frac{q(i, n_2) - q(i, n_1)}{q(i, n_1)}\right]}{\sum_i^{49} c_0(i-1)\, q(i, n_1)}$$

Thus, the relative variation of $M_1(n)$, like $M_2(n)$, between the years n_1 and n_2 is, indeed, a *weighted arithmetic average.* In the first case, it is the weighted average of the relative variations of the age-specific first marriage *rates*, the weight at age i being the rate at this age at the beginning of the period; in the second, it is the weighted average of the relative variations of the age-specific first marriage *probabilities*, the weight at age i being the product $c_0(i-1) \cdot q(i, n_1)$*: if the given structure* $c_0(i-1)$ coincides with that observed at the beginning of the period $(c_0(i-1) = c(i-1, n_1))$, the weight at age i is the same as in the first case, namely, the rate at this age at the beginning of the period. Consequently, the relative variations of M_1 and M_2 are always contained between the extreme relative variations of the age-specific rates, for the former, and probabilities, for the latter.

THE VIRTUES OF THE SYNTHETIC INDEX

Jean-Paul SARDON*

A synthetic measure is more than just an index which summarizes a set of elements: it is also a means of comparison through time and space. These conditions are fulfilled by the period fertility index TFR, which indicates the number of births that would be observed if the numbers of women of

* INED.

each age were identical, or if female birth cohorts were invariable through time and there was zero mortality until the childbearing years were over. This is a simple reduction to a standard structure, which could be replaced by other standard structures, as JLR and LT point out.

But instead of meeting the reader's expectations, by defining a more precise and more complete standard structure than the simple age structure generally used – by introducing parity and time elapsed since preceding birth – the authors deviate completely from this conception. They construct a new indicator which is naturally also based on a structure, more elaborate than the conventional age structure, but which is not defined once and for all, and changes constantly, since it depends only on the "current conditions" of fertility in the specified year.

To this annual structure, they apply the different series of birth probabilities relative to the year considered, which naturally also change over time. As a result, variations in this new indicator express the structural modifications induced by the fertility changes as well as the fertility changes themselves.

The authors claim to propose a truly 'transversal measure', but, if we leave aside their somewhat hasty assertion of the primacy of the transversal over the longitudinal approach, it seems to me that the eminently transversal nature of this measure appears above all in the fact that the behaviour of certain cohorts is applied to others.

JLR and LT think that, by using probabilities instead of rates (incidence rates or event frequencies), which do not take the cohort's past history into account, they have done away with the influence of the state of the cohort on the measure. But this assumption is not verified: a probability may be not totally independent of the proportion of persons having already experienced the event, due to a certain selection effect. Thus, to apply a probability measured from a cohort of n survivors to a cohort whose survivors are no longer n but m, means accepting an approximation which cannot be evaluated. Even if the difference were slight, multiplying such differences at every stage of the calculation might result in a total error which is not insignificant.

Moreover, if the value of a probability depends as much on 'the times' as on the state of the cohort at the beginning of the observation period, that is, on its past behaviour, then it is the very notion of 'period fertility intensity' as developed by the authors that must be rejected.

The authors' preference for synthesizing probabilities poses the problem, on top of the preceding remarks, of the comparative advantages of each of these indicators. From a longitudinal standpoint – which the authors stalwartly refuse to take – an analysis of the contradictions between cohortwise changes in rates and in probabilities reveals the incomplete, or even deceptive, nature of the information given by the probabilities. We have seen elsewhere that a rate increase always expresses the fact that the cohort is catching-up on or getting ahead of its precursor, but a probability increase is not univocal: it may mean that the cohort is reducing its 'backwardness' or, on the contrary, falling behind.

Finally, what is the significance of the annual variations in this new indicator? With a cross-sectional synthesis of incidence rates, the meaning is clear, the different elements being easily identified. In the general case of a synthesis of probabilities, their successive multiplication does not make it easy to detect the factors of change. But in the case of this composite indicator, in which several tables are integrated, it becomes impossible to identify the origin of the changes unless we have all the pieces of the puzzle.

PRINCIPAL RESULTS OF INSEE'S FAMILY SURVEY, 1990
Contribution to the debate on fertility indicators

Guy DESPLANQUES*

JLR and LT add to our array of indicators for measuring period fertility. At the same time, they challenge the exclusive use of the total fertility rate (TFR).

Should we infer that demographers have so far been devoid of imagination? L. Henry and R. Pressat, in France alone, have shown that this is not the case. Our attention has been repeatedly drawn to the dangers of over-interpreting this synthetic indicator. Nonetheless, the TFR has maintained its monopoly.

And yet the age of the computer, together with more complete data sets, from biographical surveys in particular, have opened new horizons for sophisticated calculations and greater diversity of methods. For instance, the Family Surveys(1) of 1982 and 1990 on a 1 per 50 sample of women have provided complete family formation histories.

The only information required to construct TFR is the distribution of births by mother's age, and the number of women of each age, in the specified year. Neither number of previous births nor duration since birth of last child are taken into account. Yet both these factors seem likely to play a role in the probability of giving birth during the year.

The study of first births can illustrate what is to be gained by considering some or all of these elements. The only strictly demographic information needed is, for the child, whether he/she is the firstborn or not, and for the mother, whether she has previously given birth.

---

* INSEE, INED.

(1) This survey is combined with each population census and taken by the French National Institute of Statistics (INSEE).

When we have only the former, we can calculate birth order 1 fertility rates and the first birth component of TFR, in the same way as the total first marriage rate.

The first-birth TFR is already a valuable piece of information. When it exceeds unity, the period is an exceptional one: several cohorts are, so to speak, having their first child at the same time. This overlapping generally goes together with earlier motherhood.

This was the case in the early 1960s, when the values were in the region of 1, or 0.1 point more than would have been observed in a stable fertility regime, the proportion ultimately childless being almost 12% in the cohorts born around 1935.

When we also know the number of women who have never given birth, we can calculate first-birth probabilities. These can reasonably be supposed to give a truer picture of current behaviour than fertility rates, which are sensitive to the proportion childless at beginning of year, in turn related to past behaviour.

Calculating these probabilities means considering zero-parity women, the only ones who can, in a given year, have a first child. If the first-birth TFR values were low during the late 1980s, it is to a large extent because the probabilities were low at the early reproductive ages, up to 25, and so were the rates at the higher ages, the women in these cohorts having generally already had their first child. Today, at these ages, the first-birth probabilities are not particularly low, but the proportions childless are.

The existence and the construction of several indicators raises a number of questions.

1. Is one indicator better than the rest?

The construction of an indicator reflects an – often implicit – behavioural model. In the TFR, all women of a given age, whatever their past fertility performance, are put on a same footing. The more elaborate PADTFR indicator controls for parity and duration since preceding birth. Other variables, such as women's marital status, might also, or alternatively, be taken into account. The difficulty of determining which is the best indicator increases accordingly.

Personally, I feel that it is preferable, in the absolute, to choose an indicator that eliminates the influence of the past. In the case of first births, for instance, it seems more satisfactory to use probabilities than rates.

2. Some indicators are easier to calculate than others

One of the major advantages of the TFR index is its simplicity: it is easy to calculate, at least when birth registration is adequate. The more complex indicators call for additional data. The Family Survey of 1982 provided these elements for a period of twenty years, with the biases common in retrospective survey data.

JLR and LT updated them from 1982 to 1989. For this period, they had only the distribution of children by order in marriage. They thus supposed the matrix by biological birth order and order in marriage had not changed since 1975-79, to estimate the structure by biological birth order and mother's age at birth of child.

In view of the rapid progression of non-marital births, this assumption of stability is a strong one. In addition, the Family Survey data, but also the registration data, are biased to some extent. The results of the 1990 survey indicate that the calculations made for 1982-89 slightly overestimated first births: the first-birth TFR, put at 0.77 for 1989, was 0.73 according to the survey.

This difference does not query the findings in general. But it points to the difficulty of regularly producing indicators which rest on strong assumptions. They invite criticism not so much for their complexity and number of modalities as for their over-sensitivity to the assumptions involved.

3. Too much information is not information

Communication, in the realm of statistics, is a difficult art. It is hard to explain to the general public that any form of behaviour – having children, for instance – can be measured in several different ways. There are examples every day: few people grasp the difference between the 'real' number unemployed and the number 'adjusted for seasonal fluctuations'.

This difficulty should incite us to avoid publishing systematically several indicators. But this does not mean not calculating them. Diversity reminds us that all measures are relative ones; it also broadens our understanding and thus helps us to predict the future.

«PERIOD FERTILITY INDICES» – A COMMENT

Nico KEILMAN*

The article by JLR and LT gives a clear exposition of various fertility indicators which summarize, for a given time period, more detailed rates broken down by age, parity and/or duration. I have two main points, which are meant to clarify a number of fundamental issues that JLR and LT take up more implicitly. I have no reasons to believe that the authors and I disagree strongly on these points.

* Central Bureau of Statistics, Norway.

"No model, no rate"

This is a saying which is popular among many demographers who deal with problems of analytical demography – yet its message is not always kept in mind by others. The main idea is that whatever fertility rate we compute (ranging from simple age-specific rates to very detailed ones controlling for age, parity, duration since previous event, education, labour market status, ...), there is always an underlying model, and that computing the rate in fact boils down to estimating a parameter in that model. Thus, given a set of birth data, it is not the case that, independently from the researcher's own views, there exists a set of fertility rates of any detail, which 'just need to be calculated'. There is more: the mere computation of a particular rate implies that the demographer holds a certain view about reality, expressed by the model he or she uses. This is so not only for fertility, but also for mortality, migration, nuptiality, household dynamics, etc. Let me give a simple example: an age-specific death rate. It may look a little trivial, but it illustrates many of the important points behind the saying quoted above.

Suppose we have very detailed data on deaths, which include not only information on age, but also on sex, cause of death, socioeconomic background, region of residence, etc. Computing an age-specific death rate boils down to estimating one parameter, for each age (or age group) in a Markov model with three states: alive, dead and living abroad. Individuals may jump from status 'alive' to status 'dead' – in that case, they are said to die. Jumps in the opposite direction are impossible. Immigration (abroad $\rightarrow$ alive) and emigration (alive $\rightarrow$ living abroad) are also allowed. A first-order Markov process is quite often used in such situations. In the latter process, it is assumed that at any point in time, the future behaviour of an individual only depends on his or her current situation – it is independent of his or her history (Markov assumption)[1]. Another assumption which is often made is that of independence between the three phenomena of mortality, emigration and immigration[2]. If we want to estimate, from the data, the one-year probability of dying at a certain age, we take the number of persons alive at that age, and find out how many of them are dead one year later. The ratio between these two numbers is, under certain conditions, a maximum likelihood estimate of the corresponding death probability in our simple model. One important assumption here is that all individuals contribute equally to the likelihood function – in other words, we assume homogeneity within the group of persons for whom we estimated the death probability.

Instead of the death *probability*, we could also estimate the death *rate*: the ratio between the number of deaths and the total time all individuals under

(1) Assuming a first-order Markov process is just one possibility out of many. Others include a semi-Markov process, in which future behaviour *does* depend on the past.

(2) Some demographers assume continuity as well, i.e. that immigrants have the same emigration behaviour as those initially present in state 'alive'. However, this assumption is not necessary in a first-order Markov process: it follows from the Markov assumption.

study were exposed to the risk of dying during the interval. But then we need an extra assumption in addition to the one already made, because the exposure time cannot be calculated on the basis of numbers of persons alive at the beginning and at the end of the interval. One possibility here, which is often used by demographers, is to take the simple average of initial and final population. The latter quantity would be equivalent with the total exposure time if the survivor function decreased linearly over the interval. Another possibility, which is more satisfactory from a formal point of view, is to assume a constant force of mortality (death intensity) during the interval – this is equivalent to assuming an exponential form for the survivor function(3).

In summary, when demographers 'calculate' an age-specific rate from observed data, they often assume, implicitly or explicitly, a first-order Markov process, which is based on a number of assumptions:

1. all individuals in the data who belong to the same age group have the same death rate (homogeneity);

2. the mortality behaviour depends on current conditions only, not on the past;

3. mortality is independent of disturbing phenomena (immigration and emigration);

4. the survivor function has a particular form, for instance a straight line or an exponential curve.

The first assumption implies that the researcher believes that individuals of different sex, socioeconomic background, region of residence, etc. and persons who died of different causes, all have the same death rate, because none of these covariates entered the model – only age.

This simple example illustrates a point which holds more generally when calculating occurrence-exposure rates (or probabilities, for that matter) for non-repeatable events: any rate calculation, be it for age-specific mortality, or for fertility by age and parity, or for nuptiality, migration, etc. is based on a set of assumptions, which together constitute a model. Most often the computations are carried out within the framework of a first-order Markov process, based on assumptions 1-4 above, of which the first two are the most important ones in the context of the article by JLR and LT(4). Assumption no. 1 (homogeneity) is referred to indirectly by the authors (section "The construction of the indices"). They do mention

(3) Instead of estimating the death rate directly, one could also derive its value from the estimated death probability. However, for such a computation we also need an additional assumption, for instance a linear or an exponential survivor function. The authors, when analysing fertility broken down by age plus one or more additional covariates (duration since previous birth, parity) seem to prefer to work with probabilities, instead of rates. A relative disadvantage here is that probabilities may be influenced by disturbing events. But since we work within the framework of a formal model, rates can be translated into probabilities, and vice versa.

(4) For non-repeatable events (age-specific births, without a breakdown by parity, duration, etc.), the model is that of a Poisson process, and the survivor function is constant (assumption 4). Assumptions 1-3 hold equally well in that case.

assumption no. 2 (section "A necessary assumption..."). The latter assumption is employed when calculating rates or probabilities based on a first-order Markov process (as demographers frequently do), irrespective of whether a period perspective or a cohort perspective is adopted. I got the impression that the authors state their assumption on "current conditions" (conditions du moment) for the period perspective only.

A life table is needed to derive indicators for quantum and tempo, which summarize a set of specific rates. The more detailed the rate set (only age, or also parity, duration, ...), the more complex the calculation of the summary indicators. Age-specific rates give rise to a very simple life table, with a constant l_x column. When also parity, duration, region of residence, etc. enter the rate calculation, the life table becomes more complicated, and the survivor function becomes a vector $\mathbf{l}_x$ for each age. Each type of summary indicator implies a different view on reality, a different model. In the case of computing the ISF [TFR], one believes that only age matters. Women of a certain age have the same fertility rates, irrespective of parity, duration since last birth, educational level, region of residence, etc. The ISFRAD [PADTFR] rests on the assumption that women within the same group defined by a particular combination of age, parity and duration have the same childbearing behaviour.

How far should one go in including additional covariates, provided that we have sufficient data? Statistical tests exist to determine whether an extra covariate improves the model's fit. But more important is the question whether it is easier to understand the time pattern exhibited by a certain summary indicator than the development shown by a less detailed one. For a forecaster it would be relevant to know which of these two is easier to extrapolate.

Translation

The life table discussed above facilitates combining a series of age-specific rates (possibly broken down by additional covariates) into summary indicators for quantum and tempo. Two perspectives are commonly employed. The first one is the cohort perspective, in which the set of age-specific rates stems from women born (or married, or immigrated) in the same period. These women had their births in various years. The second one is the period perspective, in which age-specific rates from one and the same period are taken from women belonging to different cohorts.

There are various pitfalls connected to the period perspective, of which I want to stress two. First, since the summary indicator is based on a set of rates *estimated for different women for each age*, it is a somewhat artificial one, as is demonstrated by the commonly used notions of *fictive* or *synthetic* cohort. A period summary indicator is thus derived from a different set of rates than a cohort summary indicator. This is true whether we use occurrence-exposure rates, or incidence rates ('taux de deuxième catégorie', 'frequencies'), or even probabilities ('quotients'). Unless age-specific rates or probabilities are constant, we will find a time pattern for period summary indicators which is different from that for cohort summary indicators. Processes of delay and catching up in cohorts are sometimes severely masked, and exaggerated in other situations. This phenomenon is known as the *translation* problem. Second, when we use incidence rates for non-repeatable events, there is an additional problem: the distribution at age x over the various states described by the vector $\mathbf{l}_x$ of survivors in the life table is not the same as that observed in reality. This means that a period summary indicator based on incidence rates for non-repeatable events does not give an accurate picture of the "true behaviour for the synthetic cohort" (sic). Hence, there are good grounds to reject the period sum of incidence rates as an indicator which reflects period quantum, as JLR and LT rightly state. But to Henry's argument, mentioned by the authors (catching up behaviour in cohorts may lead to period summary values that are too high compared to cohort values – cf. the translation problem above), one should add the fact that the incidence rates are inappropriate measures for describing fertility behaviour (cf. the second point above), and this is the reason why period sums of incidence rates may exceed 100%. (Compare also footnote 13 by JLR and LT.)

In the section entitled "Period and cohort fertility are complementary", the authors refer to work by Brass and Ryder in which attempts are made to translate cohort quantum indicators for parity-specific births into corresponding period indicators. However, it should be noted that Brass and Ryder did not use occurrence-exposure rates, but incidence rates. This leads to biased results for period quantum, cf. above. Recently, some progress was made in the analysis of the link between period and cohort indicators for parity-specific births and other non-repeatable events, focusing on occurrence-exposure rates(5).

Finally, the authors find that duration since previous birth explains little, once age and parity have been controlled for, cf. the comparison between ISFRA [PATFR] and ISFRAD [PADTFR] in section "Comparison in the framework of the general model". This is inconsistent with a number of other findings, for instance those by Martinelle for Sweden and Ní Bhrolcháin for England and Wales(6). It is unclear to me what the reasons for these differing results might be.

(5) Cf. my paper «Translation formulas for non-repeatable events», presented at the Nordic Demographic Symposium, Lund, Sweden, August 1992.

(6) See S. Martinelle, «A cohort model for analysing and projecting fertility by birth order», International Population Conference, New Delhi, September 1989. Vol. I, Liège, Belgium, IUSSP, 1989; M. Ní Bhrolcháin, «Period parity progression ratios and birth intervals in England and Wales, 1941-71: A synthetic life table analysis», *Population Studies*, 41(1), 1987.

PAST HISTORY, SYNTHETIC INDICATORS AND PERIOD FERTILITY

Maíre Ní BHROLCHÁIN*

The papers by JLR and LT on the measurement of period fertility are very much to be welcomed. I am greatly in sympathy with the attempt they make to arrive at period measures that are as free as possible of the influences of fertility in previous years. Exploration of the properties and performance of period measures such as those they examine is, in my view, an area of the highest priority in current research on fertility. There can be little doubt that currently conventional period fertility indices such as the total fertility rate are an unsatisfactory way of tracking fertility trends. In discussing the papers of Rallu and Toulemon, I shall elaborate more on points not covered by their discussion rather than dwelling on the many points of agreement. My comments are associated with ideas developed at greater length elsewhere on the cohort vs period issue, the likely primacy of period fertility and the considerations that should guide our approach to measuring period fertility[1].

Past history

To be a genuine period measure, any index of fertility in a calendar year or period should, as JLR and LT suggest, be as free as possible of the influence of fertility in previous years. Standardising for parity and duration since previous birth is an excellent way of doing this, and it is probable that age should also be a standardising factor. However, we can distinguish two kinds of influence here. First, past movements in fertility may have the effect of altering the composition of the population at risk with respect to age, parity and duration of interval since previous birth. A period measure that does not remove the compositional effects of previous fertility movements by standardising for these factors will, as a result, give a biased representation of the fertility performance of different periods relative to each other. The impact of past fertility movements on the *composition* of the population at risk is unwanted, and should be removed from period measures. Second, it is not unreasonable to suppose that there may be linkages through time between fertility in one period and that in an earlier period that are not simply the result of structural factors. Correlations of this sort, if they exist, would be different in kind from the structural, compositional effects of past fertility movements. There is no inconsistency, as I see it, between, on the one hand, the view that period fertility is of primary importance and the associated attempt to measure it

* University of Southampton, UK.

(1) M. Ní Bhrolcháin, 1992.– «Period paramount? A critique of the cohort approach to fertility», *Population and Development Review*, Vol. 18.

free of the compositional effect of previous fertility movements and, on the other, the existence of relationships between fertility in different periods. So, it is not necessary, in my view, to assume that fertility in year t is independent of fertility in years t–1, t–2, t–3, and so on: whether this is so is an empirical matter, and cannot be ruled out in principle[2]. Nevertheless, fertility in year t cannot be influenced by the (as yet unrealised) fertility of future years, since future events cannot, by definition, have a retrospective influence. It is, on the other hand, possible that fertility in year t is influenced by intentions and expectations in t regarding future fertility.

Synthetic cohort measures

JLR and LT confine their analysis to measures constructed along synthetic cohort principles: mean number of children per woman, parity progression ratios and so on. I have doubts about the value of synthetic indicators of this kind and have argued elsewhere that the synthetic cohort principle is not necessary for the measurement of period fertility[3]. That is, the case for period fertility does not depend on the validity of the synthetic cohort principle. In studying period fertility, we are concerned with the description and representation of change. There appears to be something profoundly contradictory in the convention of representing change (i.e. time trends in fertility) by an index which is interpreted through a hypothesis of unchanging fertility conditions. As I see it, we should be sceptical about the validity of representing the fertility of a period in a metric which is inappropriate to the phenomena that occur in a period. To use a mean family size as an indicator is to adopt a form of measurement that misrepresents what occurs in a period. Neither individual women nor the populations of which they are members acquire a mean number of children in a period: what happens is that fertility rates at particular ages, parities and durations are at some specified level. JLR and LT comment that in times of upheaval – such as immediately after war – period measures cannot be interpreted in terms of average number of children per woman. In my view, periods of exceptional change in fertility simply bring into sharp relief a problem that applies equally during periods of lesser change: that the synthetic cohort calculation is in principle hard to justify as a representation of period fertility at any time, if it is interpreted as representing a mean number of children per woman.

As I see it, the most defensible approach, in the strictest sense, is to represent fertility in period t by means of the disaggregated parity-, duration- and probably age-specific rates. Of course, such a representation would be awkward, requiring a large number of quantities for each period. Nevertheless, it can be argued that change in the synthetic family size indicator,

[2] Furthermore, such interdependencies need not imply a cohort basis to fertility movements.

[3] See footnote 1.

whether based on the ISFRA, ISFRD or ISFRAD calculations, can be properly interpreted only with reference to the constituent parity progression ratios, and that these in turn can be fully understood only with reference to the age-, parity- and duration-specific rates from which they are derived.

On the other hand, the profusion of rates would create a natural demand for ways of summarising the information. Summary indicators could be of the synthetic cohort type or could be constructed in a variety of other ways. The temptation to summarise by means of synthetic cohort-type quantities is strong but the logic of the argument that period fertility measures should be true to the nature of the (period) phenomena they represent would lead us to seek other ways of condensing the rates. Provided that we are free of the expectation that period measures of fertility should be of the same type as measures of cohort fertility and of the expectation that period fertility measures should fulfil the function of a reproduction rate(4), the whole question of how to summarise period rates can be thrown wide open. Viewing the parity-, duration- and age-specific rates as the fundamental building blocks for period measures of fertility, the issue of summarising these becomes a matter for empirical statistical investigation.

In my view, it would be instructive to search for alternatives to the hypothetical cohort principle for summarising period fertility rates, in their role as dependent variable. Such investigation would reveal whether synthetic summaries are, indeed, an efficient means of data reduction or whether there are other more informative methods of achieving this. A search for summary indicators should be directed at identifying the central components of fertility change, as it occurs in practice. For example, there is evidence, both in the data of JLR and LT and in earlier work, to suggest that trends in progressions of lower order (marriage, the progression to first and to second birth) behave differently from those of higher order (progression to third and later births). Further investigation could reveal whether e.g. birth probabilities at early or middle durations since previous birth behave differently or in the same way as those at later durations and so on. In my view, it is on the basis of statistical analysis of this kind that summary measures should be formulated. Such a quest would be greatly aided by, and perhaps could stimulate, attempts to provide substantive explanations of change that are more detailed and precise than those currently available to us. Indeed I believe that, as Norman Ryder has suggested, the search for better measures of temporal change in fertility should proceed hand in hand with attempts to find explanations for such change.

---

(4) I have argued elsewhere (see footnote 1) that we should distinguish clearly between the two purposes for which temporal change in fertility is measured: first, the description of trends in such a way that they are amenable to explanation (fertility as dependent variable) and second, a statement of the implication of fertility trends for future population growth prospects (fertility as dependent variable).

Formal considerations alone are almost certainly insufficient as a basis for deciding about preferred measures.

It is possible that indicators constructed according to synthetic cohort principles will be found ultimately to be adequate for the purpose of summarising the constituent period rates. If so, we need to bear in mind that a synthetic cohort indicator is merely a convenient summary device, rather than a statement about reality. We know that fertility rates are variable across time and we can therefore be sure that hypothesising fixed rates, however precisely specified, is always unrealistic. Synthetic period indicators should therefore be evaluated by how well they summarise the component period rates rather than according to how closely they conform to 'reality' in the form of real cohorts. I have argued elsewhere(5), on the basis of statistical and other considerations, that it is more probable that 'reality', in the sense of the fundamental phenomena of fertility, resides in the period rather than in the cohort domain.

A comparison with the study of mortality may be useful in this context. While a synthetic indicator, the expectation of life, is often used in summarising time-trends and cross-national differentials in mortality, it is to the components of this measure – age-specific mortality rates – that detailed investigation of the nature and causes of time-trends is usually directed. The detailed age-specific rates have not been a hindrance to mortality analysis; rather, they have been an asset. The study of age-specific rates has brought benefits in understanding the level and the causes of mortality as well as the determinants of change in mortality. It seems reasonable to suppose that similar advantages may accrue to fertility analysis from examining the detailed age-, parity- and duration-specific rates, and how these change over time. On the other hand, the mortality analogy also suggests that the use, for summary purposes, of a synthetic indicator such as the expectation of life may not be an impediment to progress. A contributory factor may be that e_0 does not play as central a role in mortality analysis as synthetic fertility measures, such as the period total fertility rate, play in fertility analysis. Overall, however, the comparison with mortality may alert us to the likelihood that attention to the detailed specific rates may be very informative.

This discussion has been concerned so far with fertility as dependent variable. Fertility has of course another aspect: from the point of view of population reproductivity, temporal change in fertility is also an independent variable – a contributor to future growth prospects and future population structure. This is a large and complex subject, and cannot be discussed at length here. We need, I believe, to re-examine the twin conventions of employing a stable population calculation to evaluate future prospects and of using a single year's or period's rates for the purpose.

(5) See footnote 1.

The arguments here are precisely the same as those applied, over forty years ago, to the use of reproduction rates. We could fruitfully reconsider the question posed by John Hajnal in 1959[6]: "In what sense is it useful to speak of measuring the reproductivity of an actual population that is not stable?" A fundamental review of this issue could result in new approaches to the measurement of fertility for the purpose of evaluating reproductivity and these could be quite different from the preferred approach to describing fertility as dependent variable. There is no *a priori* reason to expect that the measures most suitable for each purpose will coincide.

DOES ADDING MORE VARIABLES NECESSARILY IMPROVE ANALYSIS?

Patrick FESTY*

The calculation of the different synthetic measures of fertility proposed by JLR and LT is based on a more or less fine-level disaggregation of births and of the female population 'exposed to risk' (analysis). From these two elements, they calculate birth probabilities or fertility rates. They then combine these probabilities or rates by using a set of 'keys' different to those which direct observation would provide (synthesis). It is crucial to know how justifiable this change of 'keys' is for obtaining an indicator freed of the weight of past history.

Two preliminary examples

The simple, well-known example of the life table can illustrate the difficulty. In 1990, the probability of dying at age 60 is calculated on birth cohort 1930; it measures the risk in a group of people who have survived high infant mortality, no antibiotics in childhood, whose adolescence coincided with the war years, etc. Is it justifiable to combine this probability with the survivors at age 60 in the calendar year life table for 1990, which reflects infant and child protection against almost all diseases, the risks of accidental death for adolescents...? Are the survivors at age 60 in the period table, who are more numerous than those of the cohort, less fragile because of the better health care they have received, or on the contrary, less resistant because 'natural selection' has not played its role? Is it realistic to assume that the risk of dying at age 60 and the past experience of the survivors at that age are independent?

(6) J. Hajnal, 1959. «The study of fertilty and reproduction: a survey of 30 years», in *Thirty Years of Research in Human Fertility*, New York, Millbank Memorial Fund, pp. 11-37.

* INED.

Louis Henry approached the problem of fertility measures in similar terms. He was referring implicitly to the efforts of Jean Bourgeois-Pichat to construct a calendar year fertility index by combining a transversal nuptiality table with the marital fertility rates by age at marriage and duration of marriage for the same period. At the time of these studies, the fertility rates were to a large extent those of cohorts whose marriage patterns had been veered off course by the war. L. Henry showed that the many late marriages contracted in 1943 were closer, in terms of fertility, to the early marriages than had been the case previously. In fact, they would have been early marriages but for the war.

> "Intentional behaviour in marriage depends on other factors, social ones, for instance, which are independent of age at marriage. The social classes where women marry early are also those where fertility is high; marriage postponement thus led to unusually high fertility rates in the age-at-marriage group 25-29, since it was only by chance that these women were in it."

At the time, the relationship between age at marriage and fertility was, therefore, marked by this off-course behaviour, and it would have been improper to measure the weight of nuptiality on the births in the period by recalculating what the number of births would have been, had nuptiality been different: combining these fertility rates with a 'normal' marital age pattern would have been meaningless. That is why L. Henry considered it was preferable *not to disaggregate* marital births by mother's age at marriage, but only by duration of marriage.

These two examples show that the plausibility of the assumption of independence between a cohort's current and past behaviour (which is pervasive in period analysis) should always be questioned. Also, this may lead to preferring an incomplete breakdown of the phenomena studied, to avoid pursuing the 'all other things being equal' argument to the point of implausibility.

Fertility by birth order, mother's age at and duration since preceding birth

JLR and LT break down the process of family formation by birth order, just as Jean Bourgeois-Pichat broke it down into nuptiality, then marital fertility. Like the latter, the authors introduce, in their most detailed model, the age at first event and duration elapsed since this event. The probabilities calculated answer the question: what is the probability, for a mother who has had a first child at age 30, of having another the following year – at age 31 –, two years later – at age 32 – (knowing that she has not had it earlier), and so on? Using these probabilities transversally means applying them, not to the actual number of mothers concerned, but to the number obtained by combining the first-birth probabilities at ages 15-29 in the specified year.

Today, the frequency of first births has fallen among younger women and risen among older ones. First births at more mature ages are much more common in the year 1990 than in cohort 1955, for instance, which may have been used for calculating certain probabilities. The assumption of independence means postulating that second-birth behaviour after a 'late' first birth, measured in cohort 1955 where this is rare, could just as well characterize women who are much more prone to starting a family late.

Let us suppose there are three groups of women in 1990: those who continue to have their first child early, those who postpone childbearing and have it later, and those who were destined to have it later. The assumption of independence leads to postulating that the women who postpone the birth of their first child then behave like those who were destined to have it later. In the case of age at marriage discussed above, L. Henry challenged this assumption and preferred to suppose, *mutatis mutandis*, that these women kept the behaviour they would have had if their first child had been born early. Age at first birth would not be, *per se*, a relevant variable for differentiating behaviours with regard to a second birth. So it would be better to discard it and keep to second-birth probabilities by duration since first birth. This means preferring the PDTFR index in JLR and LT to PADTFR, not for reduction's sake, but because of a different assessment of the plausability of the underlying assumptions.

This prior assessment is fundamental to the analysis of current demographic trends. It is generally based on observing past time-series. To demonstrate, for instance, the inappropriateness of age at marriage for this type of fertility analysis, L. Henry considered, *inter alia*, the perturbations caused by the war, which had obliged couples to delay marrying and starting a family. But his conclusions go far beyond the critical period which served to highlight them.

Unless we postulate that there is no longitudinal logic behind behaviour at the different stages of family formation, we must resort to cohort analysis to explore the question: does postponing the first birth alter the probability of having a second child? But the data required for such an analysis are necessarily extremely fine. JLR and LT give the figure of 3,865 parameters for all birth orders. The 1 per 50 Family Survey taken with the 1982 population census provided the data needed to estimate the different synthetic measures, but not to reconstruct annual series of age-specific probabilities for measuring their changes over time. Only vital statistics would provide a body of data large enough for variations from cohort to cohort and from year to year to be meaningful.

To comprehend the demographic phenomena, we need to be aware that, say, fertility in a given year can be summarized by a variety of indicators, sometimes giving contradictory results. But, even more, we need to clearly identify the origin of their differences, since the reality of any behaviour does not lie in one measure or another, but in the plausibility of the assumptions which they imply.

PERIOD FERTILITY MEASURES
The authors' reply

Jean-Louis RALLU*, Laurent TOULEMON*

First, we should like to thank *Population*'s Editorial Board for provoking a debate around our article, and the colleagues who have kindly accepted to participate. Our reply will be in three parts, to cover their different approaches. Should period fertility be measured in terms of children per woman *during their lifetime* or *during the specified year*? How should the birth probabilities be combined? Is it always preferable to synthesize probabilities rather than incidence rates?

In what metric should period fertility be expressed?

Maíre Ní Bhrolchaín considers, as we do, that fertility behaviour should be measured from birth *probabilities*, but expresses doubts about their combination into indicators which correspond to a description of completed cohort fertility (frequency, mean age, parity progression ratios...), because "neither individual women nor the populations of which they are members acquire a mean number of children in a period".

The standardization of birth probabilities

Of the 'general period-behaviour assumption' which conventionally underlies the construction of indicators such as life expectancy at birth and Markov or semi-Markov models, as Nico Keilman has pointed out, the first two conditions apply here: 1) each group is homogeneous; 2) current probabilities are independent of the past, except for woman's age, number of children or age of lastborn, which are considered as variables relative to the mother at the specified time. To be free of this assumption, we can summarize the probabilities by using standardization methods which do not refer to the behaviour of a 'hypothetical cohort'. Fertility quantum is then measured in *children per woman per year*, or even as a relative index, instead of in *children per woman (per lifetime).*

There are three options: direct standardization (standard population), indirect standardization (standard behaviour) and regression; we prefer the latter two. Jean-Paul Sardon's suggestion – to standardize the probabilities by applying them to a standard structure by age, parity and duration since last birth – does not seem practicable for variables which are part and parcel of the behaviour studied. J. Hoem [1991, cited in our article], in an analysis of divorce risks in Sweden, demonstrates the superiority of *indirect* standardization methods, and in particuler log-linear regression on probabilities (improved indirect standardization). We fully agree with his

* INED.

conclusions, that demographic analysis should incorporate notions of modern statistical theory. The method he proposes seems to answer Maíre Ní Bhrolchaín's question on the observation of fertility performance changes in specified population groups, defined by age, parity, duration since preceding birth, or any other appropriate variable. In all cases, these standardized probabilities must be interpreted in terms of children per woman *during the specified year*.

An example of regression on birth probabilities

As an example, we take the results of logistic regression on birth probabilities by calendar year and age, performed separately for each birth order(1). In each model, the logits of the probabilities are thus estimated as sums of the calendar year and age effects. Variations in intensity are measured either on 'comparable ages, all parities combined' (i.e. age-standardized only) or 'comparable age and parity' (i.e. age- and parity-standardized).

The annual variations at each birth order are shown in the figure. The logistic variations in parity-specific probabilities in 1965 are those observed in the age-parity-year model. Between 1976 and 1989, the general fertility rates fall slightly (fertility at comparable ages, all parities combined, drops by 0.14% per year), while the probabilities rise by 0.19% per year (fertility at comparable age and parity). Here we have the difference between TFR and PATFR: TFR remains constant, despite an increasingly unfavourable parity structure.

Fertility by birth order can be compared to the parity progression ratios derived from the PATFR. For first births, the probabilities have declined regularly for the past 15 years (–1.7% a year): in fact, they have declined at the younger ages (below 28: see Toulemon, 1991, cited in our article), where many women are childless, and increased at the higher ages, where few women are still 'at risk'. For childless women of comparable age, the probability of having a child *during the year* has fallen by 2% annually. Inversely, combining birth probabilities gives the same weight to each age, and the first-birth component of PATFR remains stable: the ultimate probability of having at least one child has not changed over these 15 years (which is not shown by regression on the probabilities), and the reduction in probabilities expresses only a delay in first births. A regression including interaction between year and age would reveal this delay, but would prevent summarizing first-birth intensity by a single annual measure.

For births of orders higher than 1, the relationship between probabilities and parity progression ratios is further complicated by the timing of preceding births. The probabilities increase, but not the progression ratios:

(1) Which amounts to accepting the interactions between parity effects and age effects on the one hand, parity and year on the other. Interaction between age and year would, on the contrary, imply the advance/delay effects that we want to ignore to measure the changes in intensity by calendar year. In view of the low values of the probabilities (about 8 births per year per 100 women at risk), logistic and log-linear (proportional hazards) regressions yield very similar results. We have therefore shown the log-odds ratios as relative risks.

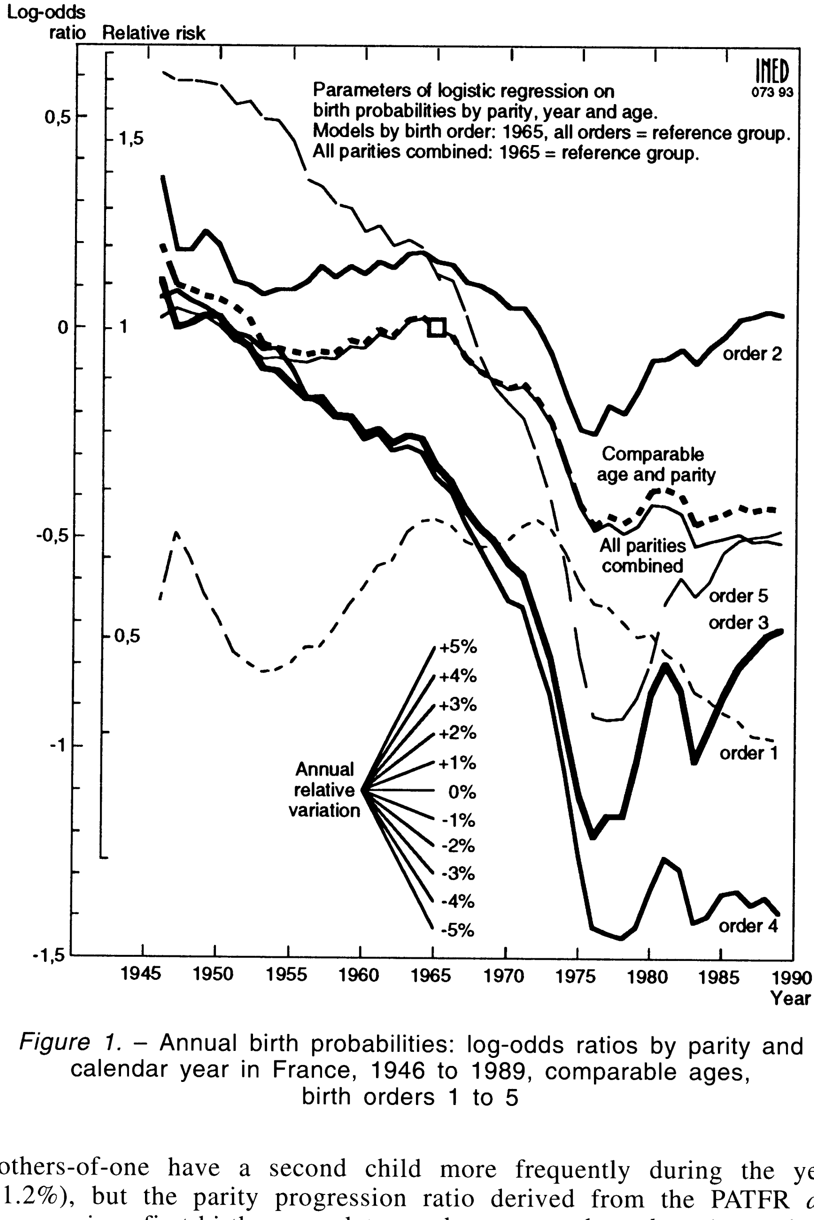

Figure 1. – Annual birth probabilities: log-odds ratios by parity and calendar year in France, 1946 to 1989, comparable ages, birth orders 1 to 5

mothers-of-one have a second child more frequently during the year (+1.2%), but the parity progression ratio derived from the PATFR *decreases*, since first births occur later, and so women have *less time* to have their second child in the 'current fertility conditions' of 1989 than in those of 1976.

For births of higher orders, the same phenomenon is observed, with the exception of fourth births: the delay in third births is not offset by an increase in order-4 probabilities (see Figures 8 to 13 in our article). Above that order, the probabilities increase sharply, because of the greater selection of parity 4+ women (who are few and from the lower social classes). The principal finding of the regression is thus that the selection effect only appears massively after the fourth birth, whereas in the past it was present after the third birth.

Regressions summarize well the instantaneous 'force of fertility' and its short-run variations, although the introduction of strongly correlated variables (age, number of children already born and duration since preceding birth, for instance) makes estimates based on 'all other things being equal' fragile and their interpretation hazardous. Regression is above all useful for comparing fertility in different population groups (defined by occupation or educational status, for instance) having 'comparable demographic variables'. But it does not yield information on the ultimate frequency corresponding to 'current conditions'. An interpretation in terms of children per woman *per lifetime* also seems useful, *and provides additional light* on the subject, since it takes into account the succession of ages and births and permits the definition of indicators comparable to those used for describing cohort fertility.

Which variables should be included in the analysis?

Can the probabilities be combined into an indicator of ultimate quantum under an assumption other than the 'general period-behaviour assumption'? And on what population groups should these probabilities be calculated?

In times of upheaval

Partick Festy reminds us of the solution proposed by Louis Henry for measuring marital fertility after the war, when couples had been obliged to postpone marrying: it was to not enter age at marriage into the analysis, fertility variations by this variable being "marked by an off-course behaviour". But Louis Henry also proposed an alternative solution [1953, chap. I, cited in our article]: to constitute groups by number of children already born, the assumption being that fertility depends essentially on this factor and on duration since preceding birth, and that the 'perturbance' in the patterns caused by the war is thus eliminated after a few years. This is what we stressed in our article: to question the independence between current probabilities and the history of the individuals is tantamount to assuming that 'unobserved heterogeneity' exists. The assumption that "there are three groups of women: those who have their first child early, those who postpone childbearing and have it later, and those who were destined to have it later" cannot be verified. But even if we accept this assumption, we must either be able to observe this heterogeneity and enter it into the construction of the indicator, or derive from it a relationship between the population structure and the probabilities, and apply to the hypothetical cohort the probabilities accordingly adjusted.

To not disaggregate is a way to re-aggregate

To describe fertility behaviour from birth probabilities, it is always justifiable to attempt to define increasingly homogeneous sub-groups. The synthesis of these probabilities can then be based on assumptions other than independence of the probabilities and the history of the individuals. But to not disaggregate for a given variable is only one possibility, and not the easiest one, except in terms of calculation. To disregard a variable amounts to considering that there would be no impact on fertility if the structure for this variable were modified (in the example given, fertility by marriage duration would be the same if war had not set marriages back). In the absence of demographic perturbations, the assumption of homogeneity in the sub-groups seems more prudent.

In this respect, the first-birth component of the conventional TFR can be presented as an adjustment of the age-specific probabilities which would suppose that, for a "same propensity", the probabilities are inversely proportional to the proportion of persons still at risk[(2)], which seems a rather strange assumption. In particular, when Gérard Calot writes that "it is even reasonable to suppose that the greater the stock, the higher the probability expressing the *same* propensity to marry as $q(i,n)$ associated with $c(i-1,n)$", this assumption, which must yet be verified, signifies that the total first marriage rate (sum of incidence rates) is even less satisfactory than the index obtained by combining the first-marriage probabilities.

We fully agree with Nico Keilman's observations on the necessity of specifying the model underlying each indicator, and on the complexity of the relations between period and cohort measures (see our article and the remarks by Maíre Ní Bhrolchaín and Guy Desplanques). He wonders why omitting 'duration since previous birth' has less impact ultimately than omitting 'age' (PADTFR is closer to PATFR than to PDTFR). Omitting a variable has no impact if: 1) fertility does not vary with this variable (once the others have been controlled for) or 2) the structure observed for this variable is, globally, neither favourable nor unfavourable to fertility. The two explanations hold here: for births of *orders 3+*, fertility varies less with duration than with age (see Figures 8 to 13); and fertility increases with duration when durations are short, then decreases when they are long (the same effects are observed for age and duration in the PADTFR model). The observed duration is longer than in the associated stable population,

---

(2) Using the notations of Gérard Calot, $q(i,n) = t(i,n)/c(i,n)$ becomes $q^*(i,n) = t(i,n)/c^*(i-1,n)$ or $q^*(i,n) = q^*(i,n) \cdot c(i-1,n)/c^*(i-1,n)$, *thus* $q^*(i,n)$ is inversely proportional to $c^*(i-1,n)$ for the age-specific marriage rates to remain constant and for the combination of probabilities to yield the same measure of intensity as the sum of incidence rates and the same mean age.

but this difference has finally no effect, except for second births (see Figure 7 and Appendix Table 4 in the original article).

Incidence rates are poor indicators

Incidence rates or probabilities?

In the case of non-renewable events, measures derived from probabilities (or occurrence-exposure rates) are based only on the population *alive* and *at risk*. The questions of homogeneity and of independence between performance and size of the sub-group are raised. But when incidence rates are used, it is *certain* that they are not independent of the group's history, since its members are extremely heterogeneous, consisting, on the one hand, of individuals who have already experienced the event considered (zero risk) and, on the other, of individuals who have not experienced the event (risk measured by the probabilities). Going from net rates to crude rates, which means excluding individuals who are zero-risk because they have died, is naturally prolonged by eliminating individuals who are zero-risk because they have already experienced the event. In the same way, for successive events, the only individuals at risk of experiencing an event are those who have already experienced the previous event. On this point, we fall in with the observations of Maíre Ní Bhrolchaín, Nico Keilman and Guy Desplanques.

The summing of incidence rates is never preferable to combining probabilities. Let us take the example of first marriages: the first-marriage probabilities fell by half between 1970 and 1990 (standardized instantaneous rates, assimilated to probabilities, proposed by Gérard Calot), but the ultimate frequency derived from these probabilities shrank by only 24%, from 0.93 to 0.71 first marriages per woman (according to the tables derived from the age-specific probabilities)[3]. In fact, a drop in probabilities (first marriages per single woman *per year*) does not produce an equivalent drop in terms of ultimate frequency (first marriages per woman *per lifetime*), because the succession of ages implies that not having married at a certain age means that one remains 'exposed to risk' at the following ages. If both measures of the periodwise decline in first marriage have a precise meaning (the probability, for a single woman, of marrying during the year has, 'at comparable ages', decreased by half in the space of twenty years; in the meantime, the ultimate probability of marrying has fallen from 0.93 to 0.71), change in the total first marriage rate (from 0.92 to 0.56 first marriages per woman, a 39% drop) is not easy to interpret, whether in terms of first marriage behaviour during the year or ultimate frequency of first marriage.

(3) The probability of never marrying has more than tripled, from 7% to 29%. A logit transformation (odds ratios) would improve the measurement of this trend.

Sudden change in the probabilities

Let us suppose, like Gérard Calot and keeping his notation, that after a long period of stability, all the marriage probabilities suddenly show a 10% leap in year n, compared to year $n-1$. He notes that the indicators M_1, based on incidence rates, and M_2, based on a standard population single, increase by 10%, while M_3 (intensity derived from the table of survivorship in the single state) only increases by 2% to 3%. Let us continue to year $n+1$. If the probabilities are the same as in year n, neither M_2 nor M_3 changes, while M_1 decreases: although nuptiality has not changed, there are fewer marriages in year $n+1$ (with constant age structure), because there are fewer single persons in year $n+1$ owing to the higher number of marriages in year n. Thus, M_1 cannot be considered satisfactory for summarizing the probabilities. In particular, if the annual variation in M_1 is always contained between the extreme variations in the incidence rates, it is not necessarily between those of the probabilities.

In this respect, the example of mortality does indeed bring "into full light" the differences between the indicators: a uniform decline in the probabilities of dying at each age is expressed by a 10% drop in M_1 and M_2, which summarize the force of mortality (or instantaneous death rate), and should therefore be expressed in deaths per person *per year*, while combining the probabilities invariably leads to a frequency of one death per person. Far from agreeing with Gérard Calot that this proves "the shortcomings of M_3", we note that the mean age at dying calculated from the incidence rates *does not change* if all the age-specific death probabilities fall by 10%, while the mean age from the table (which is, in fact, the conventional life expectancy) increases. The constancy of M_3 derived from the death probabilities means that all individuals are mortal, and that they are dying later and later. A reduction in the probabilities implies here a *postponement* of death, which M_1 and M_2 transform into a lower frequency of death.

TFR, for want of a better alternative

Fertility measures expressed as children per woman per lifetime are complementary to the study of annual birth probabilities. They are very useful for describing period fertility in concrete terms, based on the hypothetical cohort.

Among the possible summary measures, the conventional TFR only seems justified when information on birth order or parity structure is not available. The first-birth component of TFR is a poor indicator of fertility of this birth order: its variations are exaggerated, and it tends to fluctuate artificially (see our article and the above example of first marriages). Inversely, changes in the higher-order components emerge later, and are slighter, than in the corresponding probabilities. As Maíre Ní Bhrolchaín suggests, the simple relationship between the gross reproduction rate and

TFR does not give the latter any privileged role for describing fertility as dependent variable.

In conclusion

The assumption of independence between current probabilities and past history is a questionable one. We do not claim here that the 'complete' indicator, PADTFR, has no flaws (calculations are long; the amount of information required limits its application; certain probabilities must be estimated, thence imprecision and the risk of biased results; it does not take into account all the pertinent variables, as Guy Desplanques observes), or that it is acceptable without making certain assumptions. Fertility can also be measured in terms of children per woman per year within specified groups or by standardized probability methods, but in that case we lose the logical succession of ages and births. Alternatively, the probabilities can be combined by making assumptions other than homogeneity and independence from past history, although that seems the most prudent option, in the absence of 'demographic upheavals'. The construction of indicators based on parity-specific probabilities puts the conventional TFR among other possible period fertility measures, which are less influenced by the structures inherited from the past. Above all, this allows an appropriate description of period fertility by birth order.