NCEPH

Analysis of Components of Demographic Change

by

Alan Gray

March 1990

NATIONAL CENTRE FOR EPIDEMIOLOGY AND POPULATION HEALTH

WORKING PAPERS

THE AUSTRALIAN NATIONAL UNIVERSITY

NATIONAL CENTRE FOR EPIDEMIOLOGY AND POPULATION HEALTH

THE AUSTRALIAN NATIONAL UNIVERSITY

Analysis of Components of Demographic Change

by

Alan Gray

March 1990

ISBN 0 7315 0877 7
ISSN 1033-1557

WORKING PAPER NUMBER 13

National Centre for Epidemiology and Population Health
The Australian National University
G.P.O. Box 4
CANBERRA. ACT 2600, AUSTRALIA

ISSN: 1033-1557

Printed by the Australian National University
Cover by Kathy Griffiths

Introduction and theory

The techniques of demographic analysis use a variety of rates and measures, often quite simple in construction, but all requiring care when it is necessary to analyse the sources of change in their levels.

The 'components of difference' method of analysing the difference between two demographic rates, due to Kitagawa (1955, 1964), is a well-known variant of direct standardization techniques. A number of modifications and developments of Kitagawa's approach now exist (Cho and Retherford, 1973; Das Gupta, 1978, Kim and Strobino, 1984). Part of its attraction as an analytical technique is its potential for explicit attention to the question of non-additivity, also known as interaction or specificity, which Pullum (1978) notes is frequently given insufficient attention in applications of standardization. Pullum also draws attention to the analogues between standardization and multivariate techniques, and some recent work is in the direction of merging computational methods (see for example Liao, 1989). Models based on standardization can also be interpreted as a class of additive generalized linear models in which the parameters are determined by a defined structural composition, as will be seen below. Where life tables are involved in analysis of components of difference, there are special methods in common use (Pollard, 1982, 1988).

Direct standardization methods are usually used only to analyze demographic rates which are composed of other rates, specific to age or marital status or some other variable. In a more general case, suppose that a measure F can be composed in any algebraic or analytic way from a set of other measures:

$$F = F(f_a, f_b, \ldots, f_\omega).$$

If we have measurements for the ω variables f_i for two different populations (population 1 and population 2), or for the same population at two different points of time, we can express the corresponding values of the measure F in a shorthand way as:

$$F_{11\ldots1} = F(f_a(1), f_b(1), \ldots, f_\omega(1))$$

$$F_{22\ldots2} = F(f_a(2), f_b(2), \ldots, f_\omega(2)).$$

Standardization is, in this notation, a matter of calculating the function F for some mixture of the subscripts 1 and 2. In the case where there are two compositional functions f_a and f_b, there are two standardized intermediate values, namely

$$F_{12} = F(f_a(1), f_b(2))$$

and $$F_{21} = F(f_a(2), f_b(1)).$$

In the general case, there are 2^n-2 intermediate directly standardized values between $F_{11\ldots1}$ and $F_{22\ldots2}$.

Kitagawa's components of difference for the case of two compositional variables[1] is the average of two decompositions of the difference between F_{22} and F_{11}:

$$\left.\begin{aligned} F_{22}-F_{11} &= (F_{21}-F_{11}) + (F_{22}-F_{21}) \\ F_{22}-F_{11} &= (F_{22}-F_{12}) + (F_{12}-F_{11}) \end{aligned}\right) \quad \ldots(1)$$

Kitagawa's treatment of this topic refers to rate standardization models described in terms of ctors and composition. The same distinction is made by Liao (1989). The distinction between ctors and composition is omitted in the treatment in this article, since composition can be terpreted as just another variable (factor) in an algebraic sense. Thus I refer to models with 'two mpositional variables' where most other treatments of the topic, following Kitagawa, refer to ne-factor' models.

In each case, the first bracketed component on the right-hand side measures the amount of change in F due to change in the first compositional function f_a and the second component measures the amount of change in F due to change in f_b. We could also select any weighted combination of these two cases to represent the components due to each variable.

A very useful aspect of this type of analysis is that the contribution of non-linear functional interaction in the construction of F can be measured by the difference between the sizes of each component measured in the two different ways.

The functional interaction term can be made more explicit by determining another decomposition in a manner analogous to an additive linear model:

$$F_{22} = \lambda + \lambda_a + \lambda_b + \lambda_{ab}, \qquad \text{...(2)}$$

Here the main effects λ_a and λ_b are expressed as differences from F_{11}:

$$\left.\begin{aligned} \lambda &= F_{11} \\ \lambda_a &= F_{21}-F_{11} \\ \lambda_b &= F_{12}-F_{11} \\ \lambda_{ab} &= F_{22}-F_{21}-F_{12}+F_{11} \end{aligned}\right) \qquad \text{...(3)}$$

The difference between like components on the right-hand side of equation (1) is precisely λ_{ab}. The Kitagawa treatment removes this interaction term, which is due only to non-linearity in the function F, by adding it to λ_a for one decomposition and to λ_b for the other, then averaging the two resulting decompositions.[2] While the linear model is deterministic, the close correspondence with additive linear models is

---

[2]here is another linear decomposition analogous to equation (3), obtained by adding λ_{ab} to both and λ_b, and changing the sign of λ_{ab}.

evident. Moreover, the interaction term is a very useful statistic in itself: if it is zero, then all decompositions yield identical results and there is no problem in identifying the component of change due to each of the compositional variables. If the interaction term is not zero, then the cause is not interaction in the same sense as in standard multivariate analysis, for the 'dependent' variable F is only a measure constructed from the f variables. In other words F is not observed independently of its compositional variables. Interaction is merely the approximation involved in attempting a linear decomposition of a non-linear function.

The two decompositions set out in equation (1) for two compositional variables do not include non-linear functional residuals, as those of type equation (2) do. It is convenient to call decompositions without residuals *simple decompositions*, and the mean of all the simple decompositions the *Kim-Strobino mean*, because Kim and Strobino (1984) were apparently the first to set out a mean of this type, other than for the case of two compositional variables; theirs was for a special case of three compositional variables. In the general case of n compositional variables, there are n-factorial simple decompositions of the difference. Again, there is an average solution containing components from all the possible simple decompositions. For example, in the case of three compositional variables, the Kim-Strobino mean is:

Component due to first compositional variable -

$$[2(F_{211}-F_{111})+(F_{221}-F_{121})+(F_{212}-F_{112})+2(F_{222}-F_{122})]/6$$

Component due to second compositional variable -

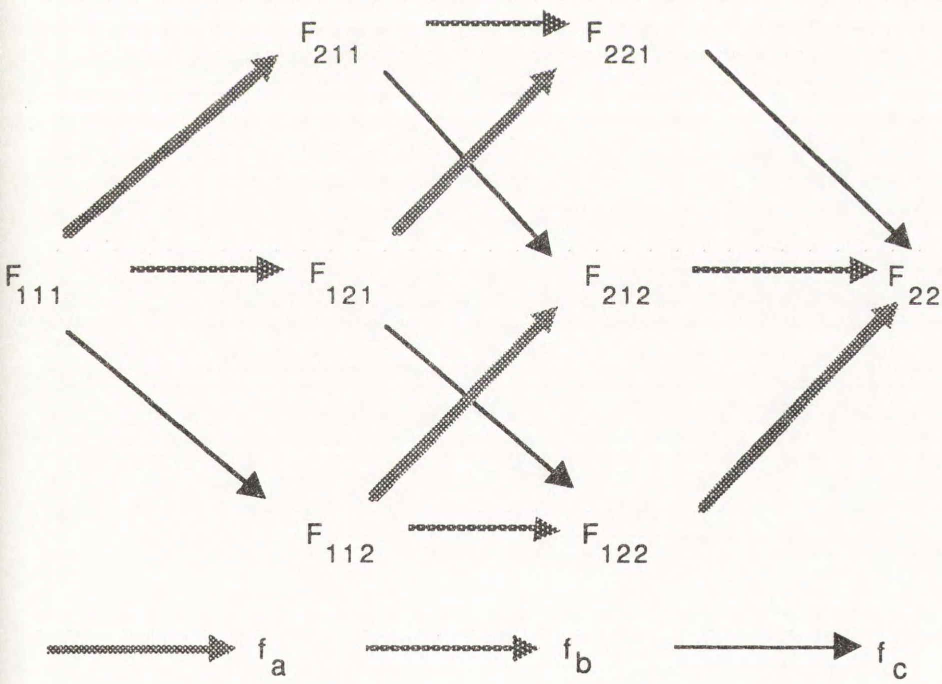
$$[2(F_{121}-F_{111})+(F_{221}-F_{211})+(F_{122}-F_{112})+2(F_{222}-F_{212})]/6$$

Component due to third compositional variable -

$$[2(F_{112}-F_{111})+(F_{212}-F_{211})+(F_{122}-F_{121})+2(F_{222}-F_{221})]/6$$

An interesting feature of the Kim-Strobino mean for three compositional variables is that some differences between standardized components are given double weight. This is because these components appear in more than one of the six (factorial 3) possible decompositions, while others appear only once. The basis of the decomposition can be visualized effectively, as in Figure 1:

Figure 1. Linear Decomposition for Three Variables



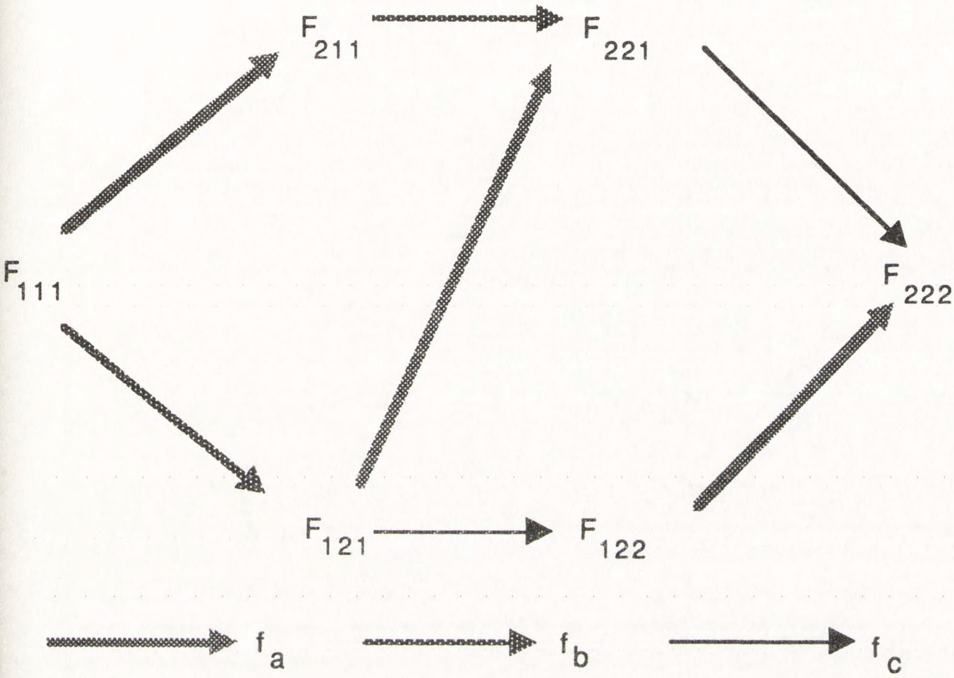
The six possible simple decompositions are represented by the six paths through the diagram from left to right, with each component arrow representing the difference between the value at its right end and the value at its left end. It is easy to see that differences in the middle of the diagram each feature in only one path, while those at the left and right of the diagram each feature in two paths. This is the source of the double weighting for some differences.

The Kim-Strobino mean becomes very complicated very quickly as the number of compositional variables is increased. There are 24 simple decompositions for the case of four variables. For five compositional variables there are 120 simple decompositions, for six 720 simple decompositions, and so on. A general method of calculating Kim-Strobino means is discussed later.

There are also many decompositions of the linear type expressed in equation (2) above. I shall restrict attention to the case analogous to equation (3), where each main effect is expressed as a difference from $F_{11...1}$. This decomposition for n variables has 2^n additive linear components, of which 2^n-n-1 are non-linear functional interaction terms.

It will be obvious that the model represented in Figure 1 assumes that the three compositional variables are independent in the sense that no judgement is made about the precedence of change in the compositional variables. There might sometimes be very good theoretical reason to suppose that one compositional variable always changes in advance of another. Suppose, for example, that f_b should, according to theory, change before f_c can possibly change, and we are dealing with a measure derived from f_a, f_b and f_c. Then some of the simple decompositions are disallowed, and the model represented in Figure 1 becomes the simpler case shown in Figure 2, with only three paths.

Figure 2. Linear Decomposition for Three Variables
-precedence of variable 2 over variable 3-



Under this model, the averaged components of change over the three allowable decompositions are:

Component due to first compositional variable -

$$[(F_{211}-F_{111})+(F_{221}-F_{121})+(F_{222}-F_{122})]/3$$

Component due to second compositional variable -

$$[2(F_{121}-F_{111}))+(F_{221}-F_{211})]/3$$

Component due to third compositional variable -

$$[(F_{122}-F_{121})+2(F_{222}-F_{221})]/3$$

Because some demographic theory (for example, demographic transition theory), would hold that fertility change should occur before mortality change in the first stages of transition, the above decomposition might be applicable in some analyses of differences in rates of natural increase. By careful attention to model specification, it might often be possible to select a reduced model incorporating precedence of change in the compositional variables, and so reduce the complexity of Kim-Strobino averages.

A cautionary note is essential here. By examining interaction terms, it might sometimes be tempting to suggest an order of precedence for change in the compositional variables so as to remove large non-linear functional interactions from the model. This approach would be invalid because the functional interactions result, as I have stressed, from the analytical approach and not the data.

Another case where one decomposition might be disallowed is the very simple case of standardized crude death rates, involving two compositional variables. The question that needs to be asked in this case is: what changes first, population age composition or age-specific death rates? This is no simple question to answer because both of these compositional variables can be considered to be related to a third hidden variable, namely the level of fertility. If we adopt a conventional view of demographic transition, then reduction of (infant and child) mortality precedes fertility decline and therefore precedes most change in population structure. In this case, a standardized crude death rate using the age-specific death rates of country A and the age composition of country B is a response to only one possible question: what would the crude death rate of country B be if it acquired the age-specific death rates of country A? A standardized death rate is most often taken to be the answer to a totally different question: what would the crude death rate of country A be if it acquired the age composition of country B? There is an invalid inversion here. However, demographic transition does not take place in discretely ordered stages and for many historical and modern populations it is no doubt much more meaningful to consider mortality and fertility as changing virtually simultaneously, in which case it is not clear if either question is valid or even that there is a question worth answer in the form of a standardized rate.

The attention given by analysts (for example Pullum, *op. cit.*) to questions of non-linear functional interaction in comparisons involving standardization techniques is obviously justified. So far we have discussed no adequate way to deal with residual interactions in the linear models (equation 2) except to remove them by averaging theoretically valid simple decompositions. Some of the other variants of decomposition techniques have more or less successfully removed these residual interactions due to non-additivity. The method of Cho and Retherford is an interesting example because it does not appear to be capable of reduction to a function of the simple decompositions discussed here. It does appear that the results obtained with the Cho-Retherford approach depend on the order in which the compositional variables are considered. The method of Das Gupta is an unsatisfying combination of the Kitagawa and Cho-Retherford decompositions. While it has already been noted that the Kim and Strobino decomposition for three compositional variables is a case of the Kim-Strobino mean, Kim and Strobino presented the method as 'hierarchical' when in fact the results obtained are independent of the order in which the compositional variables are set out.

This Kim-Strobino mean is only a convenient analytical approach if the number of compositional variables is small, and even then it may seem to be no more than a method of hiding an inherent problem of analytical method, namely functional interaction. It is actually very much more than this, in that under quite general conditions for the function F, the Kim-Strobino mean is an *exact* decomposition for the case where each compositional variable f changes linearly between $F_{11...1}$ and $F_{22...2}$ - as will be seen below.

If components of difference methods must be used, it would seem to be desirable to present, explicitly, one standard linear decomposition, including the interaction terms, so that the extent of functional interaction can be gauged by readers of the analysis. Consider a decomposition of a model involving three compositional variables:

$$F_{222} = \lambda + \lambda_a + \lambda_b + \lambda_c + \lambda_{ab} + \lambda_{ac} + \lambda_{bc} + \lambda_{abc} \qquad ...(4)$$

where

$\lambda \ \ = F_{111}$

$\lambda_a \ = F_{211} - F_{111}$

$\lambda_b \ = F_{121} - F_{111}$

$\lambda_c \ = F_{112} - F_{111}$

$\lambda_{ab} = F_{221} - F_{211} - F_{121} + F_{111}$

$\lambda_{ac} = F_{212} - F_{211} - F_{112} + F_{111}$

$\lambda_{bc} = F_{122} - F_{121} - F_{112} + F_{111}$

λ_{abc} = (most easily calculated as the remaining residual)

Presentation of all these effects makes decomposition of the components of change absolutely specific and reveals the existence of any complicating non-linear functional interactions. Matters of precedence among the compositional variables can be handled in a very straightforward way. For instance, if variable f_b always changes before variable f_c, then the appropriate decomposition becomes

$$F_{222} = \lambda + \lambda_a + \lambda_b + (\lambda_{bc} + \lambda_c) + \lambda_{ab} + (\lambda_{abc} + \lambda_{ac})$$

Here the effect of variable f_c is represented by the combined term $\lambda_{bc}+\lambda_c$ and there is also a combined interaction term and only two interaction terms in total.

The problem of non-linear functional interaction can be dealt with most adequately when good time-series data are available, with short intervals. Here we are examining an actual process of change, not attempting to fill in gaps in some hypothetical process. When series of data change gradually, it will be found that the interaction terms obtained in short interval (year-to-year) comparisons are small and sum to much smaller totals over a long period than if components of difference are examined for the two end-points of the period. It is a case of analytical integration using a fine decomposition.

Formally, the total change in a measure F, with compositional measures f_1, $f_2,\ldots,f_\omega$, between time t_1 and time t_2, can be decomposed into components as

$$\int_{t_1}^{t_2} dF(f_1,f_2,\ldots,f_\omega)$$

$$= \sum_{i=1}^{\omega} \int_{t_1}^{t_2} \frac{\partial F(f_1,f_2,\ldots,f_\omega)}{\partial f_i}\, df_i \qquad \ldots(5)$$

Each term of this sum is an exact measure of the component of difference due to change in the compositional variable concerned, but estimation is needed to calculate the component integrals. If each component integral is broken up into single-year steps, so that the interaction terms are tiny and negligible, the integral can be estimated using the main effects of a linear model. Specifically, for small steps,

$$\frac{\partial F(f_1,f_2,\ldots,f_\omega)}{\partial f_i} \approx \lambda_i = F_{11\ldots2\ldots1} - F_{11\ldots1\ldots1} \qquad \ldots(6)$$

Residual functional interaction is expected to be small but should be displayed for verification.

When the gap between t_1 and t_2 is very short, for example when the integral is broken into one-year steps, it may be reasonable to suppose that each of the component functions f_i changes linearly in that short period. Under the assumption that the compositional variables f_i change linearly, *at least in the cases where F is linear in each of a finite number of compositional variables*, the integral is the Kim-Strobino mean.[3] To see this, it is sufficient to consider the first component of the decomposition and express the integral to be calculated as:

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} F((f_1(2)-f_1(1))/n,\ f_2(2)+k(f_2(2)-f_2(1))/n,\ f_3(2)+k(f_3(2)-f_3(1))/n,\ \ldots, f_\omega(2)+k(f_\omega(2)-f_\omega(1))/n\ )$$

Evaluation of this expression yields the following results for the first few cases

[Case n=1] $F(f_1(2)-f_1(1))$

[Case n=2] $\left(\sum_i F(f_1(2)-f_1(1), f_2(i))\right)/2$

[Case n=3] $\left(\sum_{i,j} F(f_1(2)-f_1(1), f_2(i), f_3(j))+\sum_i F(f_1(2)-f_1(1), f_2(i), f_3(i))\right)/6$

[Case n=4] $\left(\sum_{i,j,k} F(f_1(2)-f_1(1), f_2(i), f_3(j), f_4(k))+2\sum_i F(f_1(2)-f_1(1), f_2(i), f_3(i), f_4(i))\right)/12$

---

his is a very general condition. It holds for all sums and products of scalar and vector variables ich do not involve powers - and so it holds for virtually all combinations of demographic rates. s not clear whether the result holds more generally, for example for measures such as ectation of life based on decomposition of rates.

The case of one compositional variable is obvious, the case of two is a generalization of Kitagawa's formula for one 'factor', the case of three is a generalization of the Kim-Strobino decomposition, and the case of four compositional variables can be shown to be the Kim-Strobino mean by constructing the rather complicated decomposition diagram for four compositional variables equivalent to Figure 1, which is for three. While this is not a formal proof for an arbitrary case, the difficulty can be circumvented by redefining the Kim-Strobino mean, in the general case, to be the limit expression given above.

Note how equation (5) makes it absolutely explicit that there are no actual, that is non-functional, interaction terms under consideration. The literature seems quite confused about whether actual interactions between variables are, or should be, measured by some of the decomposition models which retain interaction terms. In the short-step approach adopted here, there can be no confusion at all because the formulation is one where all effects are single-variable effects; if there is statistical interaction between components of change, it can be measured by time-series analysis of the small uncumulated one-year components of the main effects.

First Application: Comparison of natural increase in Malaysia and Australia, 1976

The rate of natural increase of a population can be expressed as a vector function

$$r(m,c,f) = (f-m)'.c,$$

where m is a vector of mortality rates for each age and sex, c is a vector of proportions of the population at each age and sex, and f is a vector of age-specific fertility rates of women.[4] In a comparison of the rates of natural increase of a country with reasonably high fertility and mortality, such as peninsular Malaysia in 1976, and those of a country with reasonably low mortality and fertility, such as Australia in the same year, it is usual to find that comparison of the crude rates which compose the rate of natural increase, particularly the crude death rate, is not particularly meaningful. This is because the age structures of such a pair of populations are usually completely different; it is quite often found that the crude death rate of the country with higher age-specific mortality rates is lower than the crude death rate of the low-mortality country, because the low-mortality country has a much higher proportion of its population at the higher ages where death occurs with high frequency in all populations.

Table 1 sets out mortality, population structure and fertility data for peninsular Malaysia and Australia in 1976. It is quite clear that mortality was considerably higher in Malaysia than Australia, except at young adult ages, that fertility was much higher, and that the Malaysian population was very much younger.

ıe vector m is actually (m_M m_F), where m_S is the vector specific to sex S, similarly c is (c_M c_F) and f is $_F$).

Table 1. Mortality and fertility rates, and age-sex composition, Peninsular Malaysia and Australia, 1976

	Mortality				Fertility (Females)	
	Males		Females			
	(1)	(2)	(1)	(2)	(1)	(2)
0	0.0342	0.0159	0.0263	0.0130		
1-4	0.0025	0.0008	0.0025	0.0006		
5-9	0.0011	0.0004	0.0010	0.0003		
10-14	0.0008	0.0004	0.0006	0.0003		
15-19	0.0012	0.0016	0.0008	0.0005	0.0442	0.0365
20-24	0.0017	0.0018	0.0011	0.0005	0.1946	0.1348
25-29	0.0020	0.0013	0.0013	0.0005	0.2494	0.1516
30-34	0.0025	0.0015	0.0017	0.0008	0.1922	0.0743
35-39	0.0030	0.0020	0.0023	0.0012	0.1212	0.0247
40-44	0.0050	0.0035	0.0033	0.0020	0.0448	0.0057
45-49	0.0077	0.0058	0.0050	0.0032	0.0075	0.0004
50-54	0.0126	0.0096	0.0082	0.0051		
55-59	0.0201	0.0155	0.0140	0.0077		
60-64	0.0310	0.0250	0.0211	0.0120		
65-69	0.0502	0.0392	0.0355	0.0193		
70-74	0.0683	0.0609	0.0486	0.0315		
75+	0.1216	0.1316	0.1002	0.0947		

	Population distribution			
	Males		Females	
	(1)	(2)	(1)	(2)
0	0.015	0.008	0.015	0.008
1-4	0.059	0.037	0.057	0.035
5-9	0.065	0.047	0.063	0.045
10-14	0.068	0.047	0.065	0.044
15-19	0.061	0.046	0.059	0.044
20-24	0.049	0.042	0.049	0.041
25-29	0.038	0.043	0.039	0.041
30-34	0.027	0.036	0.028	0.034
35-39	0.026	0.031	0.026	0.029
40-44	0.020	0.027	0.022	0.026
45-49	0.018	0.029	0.018	0.027
50-54	0.015	0.028	0.015	0.027
55-59	0.013	0.023	0.013	0.023
60-64	0.010	0.020	0.010	0.021
65-69	0.008	0.015	0.008	0.018
70-74	0.005	0.011	0.005	0.013
75+	0.005	0.011	0.005	0.021

(1) Peninsular Malaysia (2) Australia

The rate of natural increase in Malaysia was 2.5 per cent per annum, and in Australia 0.8 per cent per annum. Often, the difference between two such rates is expressed by simply referring to the difference between the crude death rates and the crude birth rates because the rate of natural increase is merely the difference between the crude birth rate and the crude death rate. Thus, because the difference between the crude death rates was 2 per thousand (6 per thousand in Malaysia and 8 per thousand in Australia) and the difference between the crude birth rates was -15 per thousand (32 per thousand in Malaysia and 17 per thousand in Australia), the total difference is considered to be made up of these two components. The total difference in rate of natural increase (-1.7 per cent) is made up of a negative component of -1.5 per cent contributed by fertility and another negative component of -0.2 per cent from mortality. Notice that the sign of both components is the same. This calculation is justifiable but not very helpful, because it ignores the component of difference which is due to the different age-sex compositions of the two populations.

The components of difference for three compositional variables can be calculated from the following set of unstandardized and standardized rates of natural increase:

	Source of data vector			
	Mortality (m)	Population structure (c)	Fertility (f)	Rate of natural increase
r_{111}	Malaysia	Malaysia	Malaysia	0.025404
r_{211}	Australia	Malaysia	Malaysia	0.027232
r_{121}	Malaysia	Australia	Malaysia	0.020869
r_{112}	Malaysia	Malaysia	Australia	0.011407
r_{221}	Australia	Australia	Malaysia	0.023229
r_{212}	Australia	Malaysia	Australia	0.013235
r_{122}	Malaysia	Australia	Australia	0.006053
r_{222}	Australia	Australia	Australia	0.008413

Even the briefest inspection of the standardized rates (for example, comparing r_{212} with r_{222}) shows that population structure determines a substantial component of the difference in rates of natural increase between the two populations. This is confirmed when the Kim-Strobino mean is calculated:

Components of difference (Kim-Strobino mean):

Mortality (m)	0.0021
Population structure (c)	-0.0047
Fertility (f)	-0.0144
Total	-0.0170

Expressed this way, it can be seen that the major component of the difference between the two rates of natural increase is due to fertility differences, with population structure playing a very important secondary role and mortality a completely different role than it played in the earlier calculation. The sign on the mortality component of the difference is now positive, while it was negative when the calculation was based on the crude rates. In fact, the positive sign obtained in the three-component calculation is a recognition of the lower age-specific mortality rates in Australia, because mortality is a negative component in calculation of the rate of natural increase.

Working demographers know from experience that crude birth rates may be safely compared, as reasonable measures of fertility, but that crude death rates are awfully poor measures of mortality. The results obtained here affirm that knowledge, for note that the fertility component of the Kim-Strobino mean is almost the same as the difference between the crude birth rates.

The results of the analysis are only slightly different when a complete linear model is used for the three component calculation:

Components of difference (linear model):

Mortality (m)		0.0018
Population structure (c)		-0.0045
Fertility (f)		-0.0140
Interactions -	mxc	0.0005
	mxf	..
	cxf	-0.0008
	mxcxf	..
Total		-0.0170

Notice that two of the functional interaction terms are zero by model specification, in that mortality and fertility are related additively in the formulation of the rate of natural increase and so cannot provide any functional interaction. These two terms are precisely the ones which would be subsumed by other terms if the model was restricted by requiring that fertility change precede mortality change.

Second Application: Composition of US crude death rate, by age and race, 1970 and 1985

This example has been chosen because Liao (*loc. cit.*: 722-4) has recently presented a comparison of eight methods of decomposition of differences in the crude death rate in the United States in 1970 and 1985, using race, age structure and age-race specific mortality as compositional variables. Four of the methods discussed by Liao are variants of his own of the 'purging' method of standardization, based on ideas by Clogg and Eliason (1988). The base data are presented in Liao's article and need not be repeated here.

The purging methods are said to impose contingency table or explanatory variable model structures on the data. I emphasized in the introductory section of the present article that the types of variables used in demographic analysis are usually simple algebraic constructions based on the compositional variables, and that there is little scope for considering actual interaction between the compositional variables during a period of change. I also suggested that the most appropriate way to examine actual interaction was by time-series analysis of the instantaneous (or short period) changes in the action of the compositional variables. It is not clear that purging methods are capable of identifying *functional* interaction separately from *statistical* interaction.

A comparison of the results of the various methods of decomposition set out by Liao and the results of decomposition using the methods of this article is given in Table 2. Some of the results differ from those given by Liao, and of four sets of results presented by Liao for his methods only two are given here (the partial and marginal CG and CGD methods, which Liao favours).

Table 2. Components of difference of US crude death rates, according to various methods, 1970-1985

Method	Component: Age	Race	Mortality		Other[a]
Kitagawa (1955)	1.566	0.061	-2.228		-0.081
Das Gupta (1978)	1.525	0.020	-2.228		-
Cho & Retherford (1973)	1.485	0.061	-2.228		-
Liao (1989) - partial	2.436	0.017	-3.464		0.328
- marginal	1.983	0.037	-2.999		0.296
Kim-Strobino mean	1.566	-0.020	-2.229		-
Linear model (standard case)	1.700	0.001	-2.098	ab	-0.245
				ac	-0.019
				bc	-0.026
				abc	0.004
Simple decompositions:					
(1)	1.456	-0.041	-2.098		
(2)	1.434	-0.018	-2.098		
(3)	1.700	-0.041	-2.343		
(4)	1.434	0.001	-2.117		
(5)	1.700	-0.025	-2.358		
(6)	1.674	0.001	-2.358		

[a] For the Kitagawa (1955) and Liao decompositions, this is the 'joint' effect of age and race

Source: Adapted, with additions and alterations, from Liao (1989)

The results obtained using the methods of this article evidently remain close to those obtained by the original method of Kitagawa. This is hardly surprising since the rationale of the methods is very close to the original purpose set out for decomposition methods. The most different results are those obtained with the Liao decompositions, but this is possibly because these decompositions are based on multiplicative rather than additive models.

It is noticeable that the main effect assigned to race generally has negative sign in the simple decompositions and hence in the Kim-Strobino mean, but positive sign in decompositions by other methods. It is not clear that any substantive interpretation of this effect can be adduced.

Third Application: components of change in adult life expectancy, Australia 1970 to 1980

After slow and uneven secular increase in expectation of adult life in Australia during the middle forty years of this century, there has been a comparatively spectacular increase since 1970. During the period 1970 to 1980, male expectation of life at age 15 increased by 2.65 years to 57.2 years, and the increase for women was 3.45 years to 64.4 years.[5]

Since 1970 there has also been a rapid redistribution of the proportions of adults of different marital status categories at each age and for each sex, for reasons which in some cases are continuations of long-standing trends and which in other cases represent changes of direction. Some of the reasons for change in the marital status distribution during the 1970s were: later marriage and lower rates of marriage, increased incidence of divorce, the cohort effects of historical marriage patterns, and decrease in risk of widowhood as mortality declined.

Using multiple-increment-decrement life tables, it can be shown that in 1970 the expected number of years that Australian men could expect to spend in the never-married state beyond exact age 15 was 11.7 years and in the married state 39.6 years, while by 1980 these had become 17.4 years never-married and only 34.1 years married. At 1970 rates of mortality and marital status change, women could have expected to live 8.1 years never-married beyond exact age 15 and 40.8 years married, while in 1980 these had become 14.7 years never-married and 35.4 years married. Also, the number of years expected to be spent in the divorced state increased by a multiplicative factor of about three, to 3.3 years for men and

---

ata used in this section is based on the Australian Demographic Databank - see Brown, 1979, d Krishnamoorthy and Derrick, 1983.

5.0 years for women in 1980. The decade of the 1970s was therefore a period of very rapid change in marital behaviour in Australia.

In Australia as in many other countries, there are very large differences in death rates between people who are married and people of the same age and sex who have never been married or who have lost their partners through divorce or widowhood. Why these differences exist is contentious. A plausible explanation for the fact that unmarried people have higher death rates than married people is that selection for marriage excludes the chronically ill and unfit, that the divorced got divorced partly because they were less fit, and that the widowed were themselves less fit and had married unfit partners. The conjunction of these circumstances, or any selection of them or variation on them, constitutes a 'selection' hypothesis for explaining mortality differentials by marital status. Verbrugge (1983) argues the main competing explanation, which is that the social roles provided by marriage, as well as employment and parenthood, provide protection against ill-health and mortality, mainly by influencing lifestyle.

While there is some opposition between these two hypotheses, it is possible to take the view that both types of effect operate at a certain level. Analysts interpret the evidence differently. An analysis of mortality rates of French women (Bouvier-Colle, 1983) found a high mortality pattern experienced by never-married and divorced women not in the labour force, and, age by age, a low mortality pattern containing marital status differentials for women in the labour force and married women, in the labour force or not. In explaining this result in terms of selection, Bouvier-Colle contended that never-married and divorced women not in the labour force 'owe this situation principally to their poor health'. The apparent circularity in concluding that mortality is related to ill-health illustrates an important problem in locating the existence of selection effects.

Verbrugge (*loc. cit.*), interpreting health survey data from the United States, finds similar differentials but in health status rather than mortality and shows how explanations of these differentials in terms of social roles provides an analytical framework. While she also contrasts these explanations with explanations based on selection, she does not take a conclusive stance.

The opposition between the two competing explanations is only partly inversion. What is cause in the selection hypothesis (poor health status), leading to the twin effects of poor chance of marriage and increased mortality, is an effect in the social roles hypothesis, along with increased mortality, with the cause being lifestyles associated with different marital statuses. A more comprehensive model combining the two hypotheses would result in defining a syndrome associating categories of marital status and lifestyle, on the one hand, and health status and mortality risk on the other hand.

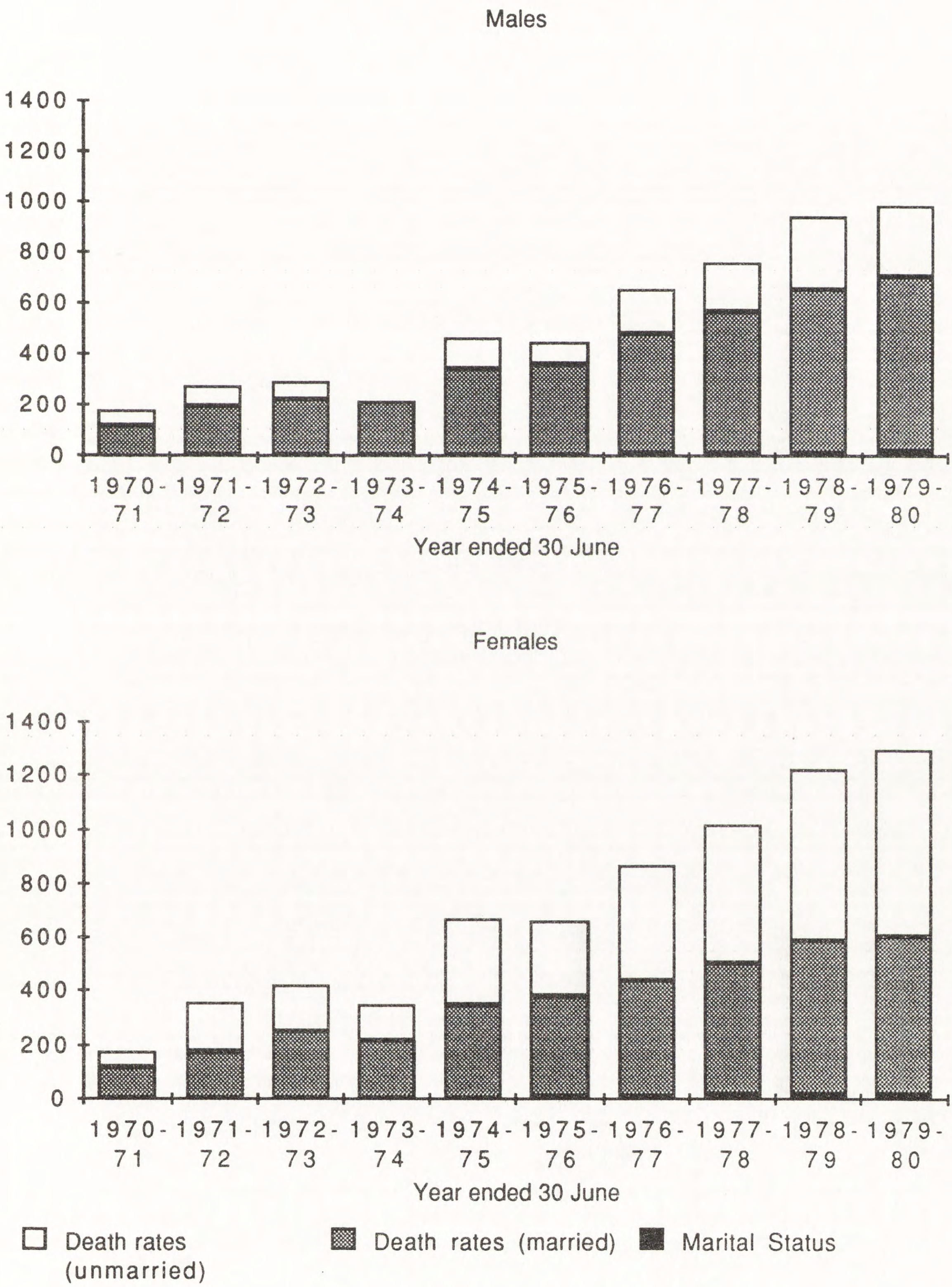
Given the rapid changes which have occurred both in marital status and expectation of adult life in Australia since 1970, it is of some interest to determine just how much of the change in expectation of adult life can possibly be attributed to change in marital status. If selection has anything to do with the matter, we should expect that a decrease in the proportion married at a particular age would result in some convergence of the age-specific death rates for married and unmarried people, because smaller proportions of 'fit' people were marrying. It is actually very difficult to determine whether such an effect exists, because of coincidental movements in age-specific death rates and chance fluctuations at the level of individual age group death rates, and because selection is unlikely to be the only influence operating on mortality differentials.

Instead of examining age group data exhaustively (and inconclusively), we may consider the increase in expectation of life at age 15 for each sex between 1970 and 1980, and the components which are due to change in the distribution by marital status at each age, age-specific death rates for married people, and age-specific death rates for categories of unmarried people.

The way in which the compositional variables here are related to expectation of life at age 15 (expectation of adult life) is that the sum of age- and marital-status-specific death rates weighted by the age-specific distribution of marital status determines each age-specific death rate, and the vector of age-specific death rates may be used to determine expectation of life. We may therefore 'standardize' by holding combinations of the three compositional variables constant from one year to another, and determining the corresponding expectation of adult life.

The somewhat surprising results of this analysis, in the form of cumulated components of the increase in adult life expectancy for each sex, are shown in Figure 3. Rapid and extensive changes in the distribution by marital status of the population had very little to do with the changes in adult life expectancy which occurred in the 1970-1980 period in Australia. Increasing proportions of people in the unmarried categories of marital status would have been expected to make some negative contribution of change in marital status to life expectancy. In fact, there were small increases in life expectancy due to change in the marital status distribution. But it is obvious that the increases in adult life expectancy are hardly sensitive to marital status redistribution.

Figure 3. Cumulative increase in life expectancy, in days, Australia, 1970 to 1980
- components of increase -



For males, most increase in adult life expectancy is due to change in age-specific death rates for married men, but for females the increase is due approximately equally to changes for both married and unmarried women. This is simply because so many women at older ages are widowed.

The components of this change for each year of the 1970-1980 period are shown in Table 3. The table also shows the sub-components of change in adult life expectancy due to change in marital status distributions below and above age 60; these have been separated because they do in fact go in different directions. Below age 60, later marriage and increased divorce contribute to a negative effect on life expectancy by placing an increasing proportion of people into categories of marital status where mortality rates are relatively high. Above age 60, there are increasing proportions of people who are married because of the cohort effects of increased rates of marriage in the past and also because decreasing death rates mean progressively fewer people suffer the loss of their spouses until later in life; both types of change work towards keeping more old people in the lower-risk category of married people.

The example is an illustration of the generality of the extensions of standardization described in the previous section of this article. The following points are worth noting. First, the function that was 'standardized' in constructing these results was no simple rate: expectation of life at age 15 was represented as a complicated function of age and sex and marital status specific mortality rates. Nevertheless, the same principles which underlie Kitagawa's components of difference of a rate can be exploited quite successfully to determine the contribution of compositional variables to change in the measure to which they contribute. Also, it has been shown that by examining year-to-year incremental change, we can successfully minimize the interaction terms which can plague analysis of the difference between end-points of a process of change.

Table 3. Components of increase in adult life expectancy (e_{15}), Australia, 1970 to 1980

Components and interaction terms[a]

	(1) Total	(1) <60[b]	(1) 60+	(2)	(3)	(1X2)	(1X3)	(2X3)	(1X2X3)	Sum
MALES:										
Increase in days of expectation of life										
1970-71	5	4	1	122	56	0	0	3	0	186
1971-72	2	0	2	70	22	0	-1	0	0	93
1972-73	-2	-3	1	28	-11	0	0	0	0	15
1973-74	4	-1	4	-13	-70	0	1	1	0	-78
1974-75	-3	-2	0	130	125	0	0	8	0	261
1975-76	-9	-11	2	13	-30	0	5	1	0	-20
1976-77	-10	-12	3	117	85	0	-1	6	0	197
1977-78	-10	-12	2	90	13	0	1	0	0	93
1978-79	-3	-5	3	85	90	0	2	1	0	176
1979-80	0	-5	5	48	-5	0	1	0	0	44
Total	-25	-47	22	690	275	0	8	20	0	967
Years	-0.07	-0.13	0.06	1.89	0.75	0.00	0.02	0.05	0.00	2.65
FEMALES:										
Increase in days of expectation of life										
1970-71	2	2	0	66	83	0	0	1	0	152
1971-72	1	-1	2	55	122	0	0	2	0	179
1972-73	0	-2	2	73	-11	0	0	0	0	62
1973-74	0	-1	1	-37	-32	0	0	0	0	-68
1974-75	1	-1	2	133	180	1	0	7	0	321
1975-76	2	-4	6	21	-37	0	2	0	0	-12
1976-77	-3	-5	2	58	150	0	-1	3	0	207
1977-78	-1	-3	2	61	88	0	2	0	0	150
1978-79	-1	-4	3	82	114	0	0	1	0	196
1979-80	-3	-4	1	15	59	2	1	0	0	74
Total	-1	-23	22	525	716	3	4	14	0	1260
Years	0.00	-0.06	0.06	1.44	1.96	0.01	0.01	0.04	0.00	3.45

[a] Component (1) is increase due to change in the marital status distribution of the population, component (2) is increase due to change in age-specific death rates for married people, component (3) is increase due to change in age-specific death rates for unmarried people. The interaction terms refer to non-linear functional interaction.

[b] Component (1) is further decomposed into components due to change below and above exact age 60.

Discussion

One emphasis in this presentation has been on a unified approach to the measurement of components of difference between demographic measures. It will have become obvious, by now, that there are many justifiable decompositions within the framework which has been set out here. The main method for choosing between decompositions should be very careful attention to a theoretical framework for demographic change. Such a framework might suggest an order of precedence of change in compositional variables. In many such cases, this kind of attention can simplify the structure of an appropriate decomposition, even though the examples of applications which have been given here do not exploit this potential for simplification because they are intended mainly for purposes of illustrating the details of calculations and interpretation.

These issues of precedence of variables, model specification and the need for a grounding in theory have analogues in regression analysis and in analysis of variance. In the deterministic models considered in this paper, they surface as issues only because decomposition of difference into additive components assumes the existence of linear relationships, and most derived demographic measures are actually not linear functions of their component measures.

In the case of measuring the components of change over time, from continuous time-series data, choice of method becomes a minor issue. This is because there is a theoretically exact method, application of which is complicated only by the inexactness of data for short discrete time periods rather than continuous time periods.

In the exact method, there are none of the functional interactions which derive from non-additivity and create specification problems for models of components of difference. While some applications will retain small residual interaction terms because of the approximations which result from data limitations, the absence of interactions in the exact method reinforces the point which has been emphasized throughout this paper, which is that measurement of change in a function derived in an analytical way from other measures may be complicated by *functional* interaction which has nothing to do with the influence of *statistical* interaction between these measures. Interaction terms always measure modelling deficiencies, primarily.

The same general point could be made by considering the variation between different decompositions. Given that there are n-factorial simple decompositions for the case of n compositional variables of a measure, a user could be tempted to calculate measures of variation between these decompositions to illustrate the precision of results obtained. This is a much more obvious case, because it should be clear that the only source of the variation is model specification, the inexactness of linear decomposition of non-linear variables. However, it is precisely this variation which is also 'interaction'.

References

Bouvier-Colle, M, 1983, 'Mortalité et activité professionelle chez les femmes', *Population*, 38, 1: 107-35

Brown HP, (updated by A.R. Hall), 1979, *Australian Demographic Databank, Vol. II: population estimates and demographic rates 1921-1976* Canberra, Department of Economics, Research School of Social Sciences, The Australian National University

Cho, LJ, Retherford, RD, 1973, 'Comparative analysis of recent fertility trends in East Asia', in International Union for the Scientific Study of Population, *Proceedings of the 17th General Conference of the IUSSP, August 1973*, Vol 2: 163-81, Liege, International Union for the Scientific Study of Population

Clogg, GC, Eliason, SR, 1988, 'A flexible approach for adjusting rates and proportions, including statistical methods for group comparisons', *American Sociological Review*, 53: 267-83

Das Gupta, P, 1978, 'A general method of decomposing a difference between two rates into several components', *Demography*, 15: 99-112

Kim, YJ, Strobino, DM, 1984, 'Decomposition of the difference between two rates into several components', *Demography*, 21: 361-72

Kitagawa, EM, 1955, 'Components of a difference between two rates', *Journal of the American Statistical Association*, 50: 1168-74

Kitagawa, EM, 1964, 'Standardized comparisons in population research', *Demography*, 1,1: 296-315 (1964)

Krishnamoorthy, S,. Derrick, B,1983, *Australian Demographic Databank, Vol. III: recorded vital statistics, population estimates and demographic rates 1976-1981* Canberra, Department of Demography, Research School of Social Sciences, The Australian National University

Liao, TF, 1989, 'A flexible approach for the decomposition of rate differences', *Demography*, 26, 4: 717-26

Pollard, JH, 1982, 'The expectation of life and its relationship to mortality', *Journal of the Institute of Actuaries*, 109: 225-240

Pollard, JH, 1988, 'On the decomposition of changes in expectation of life and differentials in life expectancy', *Demography*, 25: 265-76

Pullum, TW, 1978, *Standardization*, World Fertility Survey Technical Bulletin No.3/Tech.597, Voorburg, International Statistical Institute

Verbrugge, LM, 1983, 'The social roles of the sexes and their relative health and mortality', in A.D. Lopez and L.R. Ruzicka (eds.), *Sex Differentials in Mortality: trends, determinants and consequences*, pp. 221-45 Canberra, Department of Demography, Research School of Social Sciences, The Australian National University

National Centre for Epidemiology and Population Health
The Australian National University
G.P.O. Box 4
CANBERRA. ACT 2600, AUSTRALIA

ISSN: 1033-1557

Printed by the Australian National University
Cover by Kathy Griffiths

NATIONAL CENTRE FOR EPIDEMIOLOGY AND POPULATION HEALTH

WORKING PAPERS SERIES **ISSN 1033-1557**

WORKING PAPER NUMBER 1
Noor A. Khalidi *Aboriginal Mortality in Central Australia, 1975-77 to 1985-86: A Comparative Analysis of Levels and Trends*
April 1989 ISBN 0 7315 0801 7

WORKING PAPER NUMBER 2
Alan Gray *Discovering Determinants of Australian Aboriginal Population Health*
May 1989 ISBN 0 7315 0629 4

WORKING PAPER NUMBER 3
P.J. Solomon, J.A. Doust and S. R. Wilson *Predicting the course of AIDS in Australia and evaluating the effect of AZT: A first report*
June 1989 ISBN 0 7315 0660 X

WORKING PAPER NUMBER 4
John Deeble *The Medicover Proposals*
August 1989 ISBN 0 7315 0737 1

WORKING PAPER NUMBER 5
J.R.G. Butler *Intergovernmental Relations and Health Care*
August 1989 ISBN 0 7315 0740 1

WORKING PAPER NUMBER 6
Alan Gray *Aboriginal Fertility: trends and prospects*
October 1989 ISBN 0 7315 0809 2

WORKING PAPER NUMBER 7
Dr John McCallum *The Public versus the Policies: the Ethical Basis of Australian Retirement Income Policy*
November 1989 ISBN 0 7315 0798 3

WORKING PAPER NUMBER 8
Noor A. Khalidi *Aboriginal Fertility in Central Australia*
November 1989 ISBN 0 7315 0825 4

WORKING PAPER NUMBER 9
Dr John McCallum *The Dynamics of Community Involvement in Old Age: the Syndrome of Underuse*
November 1989 ISBN 0 7315 0826

WORKING PAPER NUMBER 10
Dr John McCallum *Noncontributory Pensions for Developing Countries: Rehabilitating an Old Idea*
December 1989 ISBN 0 7315 0838 6

WORKING PAPER NUMBER 11
M. Mackisack, A.J. Dobson & C.R. Heathcote *Predicting the Prevalence of a Disease in a Cohort at Risk*
December 1989 ISBN 0 7315 0839 4

WORKING PAPER NUMBER 12
C.R. Heathcote, C. Keogh, T.J. O'Neill *The Changing Pattern of Coronary Heart Disease in Australia*
December 1989 ISBN 0 7315 0840 8

Copies of Working Papers still in print may be obtained from: The Publications Officer, NCEPH, Australian National University, GPO Box 4, CANBERRA A.C.T. 2601, AUSTRALIA

ISSN 1033-1557