Int. J. Epidemiol. Advance Access published May 20, 2004

International Journal of Epidemiology 2004;**33**:1–10
DOI: 10.1093/ije/dyh156

Income and mortality: the shape of the association and confounding New Zealand Census-Mortality Study, 1981–1999

Tony Blakely[1], Ichiro Kawachi,[2] June Atkinson[1] and Jackie Fawcett[1]

Accepted 26 January 2004

Objective To determine the shape of the income–mortality association, before and after adjusting for confounding by other socioeconomic variables.

Methods Poisson regression analyses were conducted on 11.7 million years of follow-up of 25–59 year old New Zealand census respondents spanning four separate cohort studies (1981–1984, 1986–1989, 1991–1994, and 1996–1999).

Results Mortality among low-income people was approximately two times that among high-income people. Adjustment for potential socioeconomic confounders (marital status, education, car access, and neighbourhood socioeconomic deprivation) halved the strength of the income–mortality association, but did not appreciably change the shape of the association. Further adjustment for labour force status largely removed the income–mortality association. The association of non-transformed income with mortality was non-linear, with a flattening out of the slope at higher incomes. Both the logarithm and rank of income appeared to have a better linear fit with the mortality rate, although the association of mortality with the logarithm of income flattened out notably at low incomes.

Conclusions Much, but not all, of the crude association of income with mortality could be due to confounding. Adjusting income–mortality associations for labour force status (also a proxy for health status) is problematic: on the one hand, it over-adjusts the association as poor health will be on the pathway from income to mortality; on the other hand, it appropriately adjusts for both confounding by labour force status and reverse causation whereby income changes as a result of poor health. Both logarithmic and rank transformations of income have a reasonable linear fit with income.

Keywords Income, mortality, shape, rank, inequality, confounding, comparisons over time

Income is strongly associated with mortality,[1–8] although not all of the crude association of income with health is causal.[9] Income is associated with other socioeconomic factors, lifestyles, personality type, and other variables that also have an independent association with health. Disentangling these potentially confounding associations is difficult. Many of these associated variables are likely to reflect both mechanisms on the causal pathway from income to health (e.g. income enables a healthy diet) and mechanisms not on the causal pathway (e.g. healthy diet is a function of variables other than just income). Controlling for such variables will, in principle at least,[10–13] appropriately remove some aspect of the confounding, but also 'over-adjust' for some aspect of the causal pathway from income to health. A particularly problematic variable in the context of income is labour force status. It is undoubtedly an important determinant of income, and highly likely to have an association with health that is independent of income. However, it is also a proxy for health status—people in poor health tend not to be active in the labour force.

Assuming that at least a portion of the association between income and mortality is causal, the health effects of income redistribution will depend, among other things, on the shape of

[1] Department of Public Health, Wellington School of Medicine, University of Otago, PO Box 7343, Wellington, New Zealand. E-mail: tblakely@wnmeds.ac.nz

[2] Department of Health and Social Behaviour, and the Harvard Centre for Health and Society, Harvard School of Public Health, 617 Huntington Ave, MA 02115, USA. E-mail: ichiro.kawachi@channing.harvard.edu

the income–mortality association. If the shape is linear on the additive scale (i.e. each extra dollar buys an equivalent additional quantum of mortality risk reduction) then income redistribution will reduce social inequalities in mortality, but it will not affect average mortality rates. Income redistribution will result in a zero-sum situation, with health gains among the poor equalling health losses among the wealthy (assuming an income transfer from the rich to the poor). Alternatively, if each extra dollar buys a little less protection against death, then redistribution of income away from the rich to the poor will not only reduce social inequalities in mortality but also reduce overall mortality rates. This positive-sum situation arises because the increased mortality risk among the rich resulting from the income transfers is more than offset by the decreased mortality risk among the poor. Most research supports the diminishing returns of additional income on mortality risk.[1,2,14,15] For example, Backlund *et al*. found that the association of the *logarithm* of income with mortality was almost linear. However, others have contested that mortality is linearly associated with money income (with no transformation). For example, Martikainen *et al*. concluded that the association of income and mortality in Finland was linear during the early 1990s—a period of major economic recession and rising unemployment.[3] Martikainen *et al*. point to this linearity as suggestive of the Nordic welfare state protecting against the threshold effects of poverty that produces a non-linear association of income and health in other countries.

The shape of the association between income and mortality is also of methodological interest to social epidemiologists summarizing the strength of the mortality gradients for cross-regional and times series comparisons. The slope and relative indices of inequality (SII and RII)[16] are common and parsimonious summary measures that are increasingly used.[17–19] These two measures assume that there is a linear association of mortality (or other health outcomes) with the *ranking* of individuals by the socioeconomic factor in question. A case has also been made for the theoretical importance of income rank over and above actual income as a determinant of health.[20]

The first objective of this paper is to determine the linearity (or not) of the income–mortality association for each of the three specifications of income identified above: (1) absolute dollars, (2) natural logarithm of income (diminishing marginal returns) and (3) ranking by income. A second objective of this paper is to examine potential confounding of the income–mortality association by other socioeconomic variables.

Method

We used a series of four cohort studies of all New Zealand census respondents (1981, 1986, 1991, and 1996 censuses) formed by record linkage of each census to 3 years of subsequent mortality data.[21,22] Some 1,469 484, 1,614 678, 1,718 106, and 1,913 544 people aged 25–59 years completed each of the four successive censuses. Of the 12 792, 12 423, 11 646, and 11 940 mortality records for people aged 25–59 years at death and dying within 3 years of census night, 68.3%, 69.4%, 71.8%, and 71.8% were linked back to the respective census. The proportion of mortality records linked back to a census record varied by age and ethnicity, but within strata of age by ethnicity (i.e. variables controlled for in all analyses in this paper) there was relatively little linkage bias by socioeconomic position.[22,23] Nevertheless, to adjust for any possible remaining linkage bias by socioeconomic position, weights were used in all regression analyses in this paper. The weights were calculated, within strata, as the inverse of the probability of linkage of a mortality record to a census record. Strata were based on sex, age, ethnicity, rurality, cause of death, and small area socioeconomic deprivation according to mortality data.[24]

We calculated the equivalized real household income from individual census record data as follows. Total personal income was collected in 24, 16, 13, and13 bin-categories for each of the four censuses, respectively. We assigned each individual the median household income for the same income band from the corresponding New Zealand Household Economic Survey for the latter three censuses and mid-point income (and Pareto estimate for top category) for the 1981 census. The total household income was then equivalized for household size and economies of scale using the New Zealand-specific Jensen Index[25] (ref. 23, pp. 33–34). The equivalized household income on the 1981, 1986, and 1991 census cohort data was further adjusted to 1996 real dollars using the consumer price index.

Demographic covariates were categorized as follows: 5-year age groups; three-category ethnicity (Mäori, Pacific, non-Mäori non-Pacific); and currently married versus not currently married. Educational qualification was coded as post-school, school, and nil. Household car access was categorized as zero, one, and two or more cars. Labour force status was categorized as employed, unemployed (but looking for work and available for work), and non-active. Quintile of small area socioeconomic deprivation was assigned to individual census records using either the 1991 or 1996 variants of a New Zealand-specific index.[26,27]

We excluded the first 6 months of follow-up for each of the four census cohorts to reduce the possibility of health selection effects arising from a decline in income preceding death. Further, we used only census records with complete data for the above variables, resulting in 79.2%, 81.5%, 84.7%, and 84.0%, respectively for the four cohorts, of the total eligible follow-up being available for analysis. The majority of exclusions were due to missing total household income data. Sensitivity analyses published elsewhere suggested little likely selection bias due to this necessary restriction.[23]

Rate ratios for all-cause mortality by income were calculated using Poisson regression on the individual-level cohort data. Income was initially modelled as a 10-level categorical variable.

We used scatter plots to examine the shape of the income–mortality curve. The y-axes corresponded to the rate ratio for each income category, and the x-axes corresponded to the 'midpoint' of the income category. This midpoint was plotted in three ways. First, we used the median dollar income of each of the 10 income categories. Second, we used the natural log of the median dollar income of each of the categories. Third, a rank-transformation value (or ridit score) between 0 (person with lowest household income) and 1 (person with highest household income) was assigned.

To supplement a visual interpretation of these plots, we calculated change of deviance statistics using a second round of Poisson regression models. The only difference for this second round was that the income values were no longer modelled as a 10-categorical variable, but rather modelled as the actual equivalized household income dollar value, logarithm of these dollar values, or ridit scores based on these dollar income values within the population. (That is, income was modelled separately

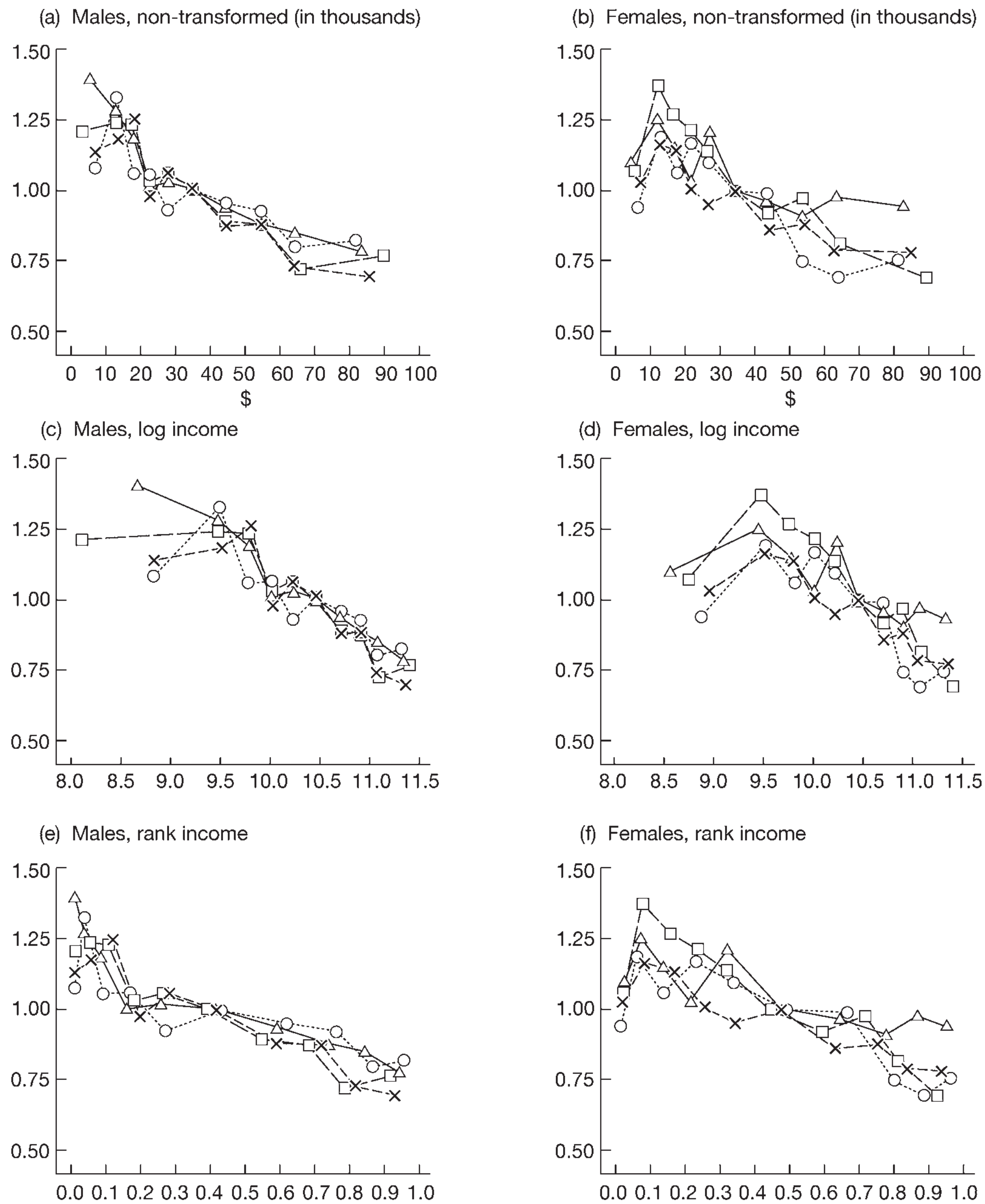

Figure 1 Rate ratios of mortality by cohort (model 2 in Table 3) plotted by different transformations of equivalized real household income on the x-axis: non-transformed (i.e. in dollar units); log transformed; ranked

as three different continuous variables, rather than as nine dummy variables that generated categorical rate ratios which we could then plot using different transformations of the income variable on the y-axes in Figure 1.) To determine the best fitting option for the three transformations of income, we compared the change in deviance statistic for the common baseline model that did not include the transformed income variable. Because of concerns about the reliability of very high or very low income recorded on census data that would distort regression coefficients for continuous variable transformation of income, we discarded all people with an income ⩽2.5th percentile or ⩾97.5th percentile for these latter Poisson regression analyses. Note that the regression models used to calculate the deviance statistics force a multiplicative association of the mortality rate with the three income transformations specified as continuous variables. Thus the models are equivalent (visually) to a log-scale for the rate ratios on the y-axes in Figure 1, a loss of comparability that we will expand on in the Discussion.

Results

Table 1 presents the distribution and rate ratios of mortality by equivalized household income for males and females in each of the four census cohorts. Table 2 shows the age and ethnicity

Table 1 Person time and deaths (linked and weighted) by equivalized household income among 25–59 year old deaths during four census-cohort periods: 1981–1984, 1986–1989, 1991–1994, 1996–1999

Equivalized real household income	1981–1984			1986–1989			1991–1994			1996–1999		
	Person time	Linked deaths	Weighted deaths	Person time	Linked deaths	Weighted deaths	Person time	Linked deaths	Weighted deaths	Person time	Linked deaths	Weighted deaths
Males												
⩾$70 000	142 974	291	417	127 278	240	336	221 493	306	408	278 772	369	504
$60 000–$69 999	97 296	222	318	117 612	213	303	126 384	177	252	149 472	192	267
$50 000–$59 999	156 336	393	561	167 880	378	543	162 330	297	420	191 736	312	441
$40 000–$49 999	199 131	531	765	231 519	567	813	234 186	429	615	245 037	408	585
$30 000–$39 999	245 739	639	939	285 423	708	1 035	280 197	612	876	267 930	510	747
$25 000–$29 999	141 561	324	492	150 195	309	450	140 685	327	477	153 042	324	468
$20 000–$24 999	109 899	243	372	138 219	294	447	116 298	225	345	112 731	219	321
$15 000–$19 999	76 644	198	309	89 955	201	297	111 378	354	507	111 408	315	468
$10 000–$14 999	40 0774	144	216	55 413	186	282	91 362	315	471	80 334	264	384
<$10 000	25 947	99	153	27 942	75	111	34 113	93	141	44 205	123	183
Total	1 235 601	3078	4542	1 391 436	3171	4617	1 518 426	3135	4512	1 634 667	3 036	4368
Females												
⩾$70 000	120 975	168	237	105 945	111	147	197 541	189	240	258 813	201	252
$60 000–$69 999	85 710	129	180	100 500	96	129	120 228	126	156	140 817	138	171
$50 000–$59 999	136 746	204	288	145 623	162	219	149 742	177	231	183 177	225	282
$40 000–$49 999	189 078	291	411	213 240	330	450	225 582	273	345	238 116	279	348
$30 000–$39 999	243 336	360	504	274 749	414	567	276 960	405	525	272 442	351	447
$25 000–$29 999	143 199	225	324	153 273	225	318	143 166	198	261	159 678	249	321
$20 000–$24 999	118 845	162	237	147 771	219	312	131 193	180	240	126 954	192	246
$15 000–$19 999	91 164	150	219	113 022	168	237	136 110	267	360	144 975	276	366
$10 000–$14 999	65 013	162	234	97 110	174	249	144 510	264	366	125 268	240	324
<$10 000	57 402	90	138	37 824	48	72	55 119	75	99	75 183	102	135
Total	1 251 465	1938	2763	1 389 057	1947	2700	1 580 151	2154	2823	1 725 423	2253	2892

Raw numbers are random rounded to the nearest multiple of three as per SNZ protocol.

Table 2 Age and ethnicity adjusted rate ratios (95% CI) of mortality by equivalized household income among 25–59 year old deaths during four census-cohort periods: 1981–1984, 1986–1989, 1991–1994, 1996–1999

Income	1981–1984	1986–1989	1991–1994	1996–1999
Males				
⩾$70 000	0.70 (0.61, 0.81)	0.70 (0.61, 0.81)	0.60 (0.52, 0.69)	0.64 (0.56, 0.73)
$60 000–$69 999	0.79 (0.70, 0.90)	0.72 (0.64, 0.82)	0.64 (0.56, 0.74)	0.65 (0.56, 0.74)
$50 000–$59 999	0.86 (0.76, 0.97)	0.87 (0.77, 0.98)	0.82 (0.72, 0.95)	0.81 (0.71, 0.94)
$40 000–$49 999	0.92 (0.82, 1.03)	0.91 (0.81, 1.02)	0.84 (0.74, 0.95)	0.85 (0.75, 0.97)
$30 000–$39 999	1	1	1	1
$25 000–$29 999	1.05 (0.92, 1.19)	0.93 (0.81, 1.06)	1.12 (0.98, 1.28)	1.12 (0.97, 1.28)
$20 000–$24 999	1.04 (0.90, 1.21)	1.08 (0.94, 1.23)	1.03 (0.88, 1.19)	1.09 (0.93, 1.27)
$15 000–$19 999	1.26 (1.07, 1.47)	1.09 (0.93, 1.27)	1.43 (1.25, 1.63)	1.43 (1.25, 1.64)
$10 000–$14 999	1.41 (1.17, 1.68)	1.49 (1.27, 1.75)	1.48 (1.29, 1.69)	1.57 (1.35, 1.82)
<$10 000	1.53 (1.24, 1.90)	1.15 (0.91, 1.46)	1.28 (1.03, 1.59)	1.40 (1.15, 1.70)
Females				
⩾$70 000	0.83 (0.69, 1.00)	0.64 (0.52, 0.79)	0.67 (0.56, 0.79)	0.60 (0.50, 0.71)
$60 000–$69 999	0.90 (0.76, 1.07)	0.62 (0.52, 0.76)	0.71 (0.59, 0.85)	0.74 (0.62, 0.89)
$50 000–$59 999	0.87 (0.73, 1.03)	0.70 (0.58, 0.84)	0.82 (0.69, 0.98)	0.91 (0.77, 1.08)
$40 000–$49 999	0.94 (0.81, 1.10)	0.95 (0.83, 1.10)	0.82 (0.71, 0.96)	0.89 (0.76, 1.04)
$30 000–$39 999	1	1	1	1
$25 000–$29 999	1.25 (1.05, 1.47)	1.12 (0.96, 1.32)	0.99 (0.83, 1.17)	1.19 (1.01, 1.39)
$20 000–$24 999	1.07 (0.89, 1.29)	1.22 (1.04, 1.44)	1.07 (0.90, 1.27)	1.29 (1.08, 1.54)
$15 000–$19 999	1.26 (1.04, 1.52)	1.13 (0.95, 1.35)	1.26 (1.08, 1.46)	1.43 (1.22, 1.67)
$10 000–$14 999	1.45 (1.21, 1.75)	1.38 (1.16, 1.64)	1.37 (1.18, 1.60)	1.66 (1.41, 1.95)
<$10 000	1.27 (1.01, 1.60)	1.03 (0.76, 1.39)	1.17 (0.92, 1.50)	1.27 (1.02, 1.58)

The rate ratios are from a Poisson regression model with age in 5-year age groups and ethnicity specified as Maori, Pacific Island, and non-Maori non-Pacific.

adjusted rate ratios. As the number of deaths and census respondents tended to be greatest in the $30 000–$39 999 category, we treated this category as the reference category. Mortality in the lowest (<$10 000) income category was approximately twice that in the highest income category (⩾$70 000) for both sexes and all four cohorts. Whilst CI were wide for the <$10 000 group, in all but one instance (males 1981–1984) the rate ratios for the second lowest income category ($10 000–$14 999) were greater than those for the <$10 000 group. There was also an apparent tendency for the income–mortality association to strengthen from 1981–1984 to 1996–1999.

Confounding of the income–mortality association

Compared with the age- and ethnicity-adjusted model in Table 2, adjusting further for education and marital status (i.e. potential confounders of the income–mortality association; model 1 in Table 3) reduced the strength of the income–mortality association by about a quarter. The second model in Table 3 further adjusts for car access and neighbourhood socioeconomic deprivation (i.e. variables that may both be confounders and mediators of the income–mortality association), halving the strength of the association compared to the age- and ethnicity-adjusted model. The third model further adjusts for labour force status, which largely removes the income–mortality association. The fourth model restricts the second model analysis to just those in active employment, attempting to overcome the problem of non-activity in the labour force being a proxy for poor health. Again, the association of income with mortality is mostly reduced to the null.

Shape of the income–mortality association

Figure 1 shows the shape of the income–mortality association by plotting the categorical rate ratios for model 2 in Table 3 according to the three transformations of income on the x-axis. (That is the rate ratios are adjusted for potential confounders other than labour force status.) A visual inspection of Figures 1a and 1b suggests that the slope of decline in mortality risk with increasing income on a dollar or absolute scale flattens out above $30 000. The plots where the income x-axis is on a log-scale (Figures 1c and 1d) are mostly linear except for the flattening (or even change in direction) of the slope among low-income people that is made more apparent by the log-scale stretching out the lower income end of the scale. However, only 2–4% of person time was in the lowest income group (<$10 000, Table 1). The plots by rank-transformed income (Figures 1e and 1f) de-emphasize the exaggerated flattening (or reversal) at low incomes by compressing these observations that contain a small proportion of the total person time. Apart from a possible steepening of the income–mortality association beneath the 20th percentile of males ranked by income, the rank-transformed plots are reasonably consistent with a linear fit.

Table 4 shows the reduction in the deviance statistics for the three alternative transformations of income specified as continuous variables in a second round of Poisson regression models. The

Table 3 Multivariable adjusted rate ratios of mortality by equivalized household income among 25–59 year old deaths during four census-cohort periods: 1981–1984, 1986–1989, 1991–1994, 1996–1999

Model covariates	Model 1. Age, ethnicity, marital status, and education				Model 2. Model 1 plus car access and small area socioeconomic deprivation				Model 3. Model 2 plus labour force status				Model 4. Model 2 restricted to employed			
	1981–1984	1986–1989	1991–1994	1996–1999	1981–1984	1986–1989	1991–1994	1996–1999	1981–1984	1986–1989	1991–1994	1996–1999	1981–1984	1986–1989	1991–1994	1996–1999
Males																
⩾$70 000	0.73	0.74	0.62	0.69	0.78	0.82	0.69	0.76	0.81	0.84	0.73	0.81	0.81	0.86	0.74	0.84
$60 000–$69 999	0.82	0.75	0.68	0.67	0.85	0.80	0.73	0.72	0.88	0.82	0.77	0.77	0.87	0.86	0.74	0.80
$50 000–$59 999	0.86	0.89	0.83	0.83	0.88	0.92	0.88	0.87	0.91	0.95	0.92	0.92	0.90	0.98	0.94	0.95
$40 000–$49 999	0.93	0.93	0.85	0.87	0.94	0.95	0.88	0.89	0.96	0.97	0.91	0.92	0.95	0.98	0.88	0.92
$30 000–$39 999	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$25 000–$29 999	1.04	0.94	1.09	1.09	1.03	0.92	1.06	1.06	1.00	0.88	1.00	0.99	1.00	0.89	1.07	0.95
$20 000–$24 999	1.03	1.09	1.02	1.08	1.01	1.06	0.98	1.03	0.96	0.97	0.87	0.91	1.01	0.99	0.81	0.86
$15 000–$19 999	1.23	1.09	1.36	1.35	1.19	1.06	1.25	1.23	1.04	0.88	1.02	0.98	0.97	0.87	0.95	0.95
$10 000–$14 999	1.34	1.41	1.34	1.42	1.28	1.33	1.18	1.24	0.90	0.92	0.88	0.93	0.79	0.94	0.82	1.11
<$10 000	1.47	1.10	1.23	1.34	1.40	1.08	1.13	1.21	1.01	0.83	0.89	0.92	0.81	1.06	1.11	1.16
Females																
⩾$70 000	0.88	0.68	0.72	0.65	0.94	0.75	0.78	0.69	1.03	0.79	0.84	0.73	1.14	0.90	0.91	0.71
$60 000–$69 999	0.93	0.65	0.75	0.78	0.98	0.69	0.79	0.81	1.08	0.74	0.86	0.90	1.11	0.74	0.92	0.82
$50 000–$59 999	0.88	0.72	0.85	0.94	0.91	0.75	0.88	0.97	1.00	0.79	0.97	1.04	1.07	0.85	0.98	0.99
$40 000–$49 999	0.95	0.97	0.84	0.90	0.97	0.99	0.86	0.92	1.02	1.02	0.91	0.96	1.11	1.00	0.93	0.95
$30 000–$39 999	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
$25 000–$29 999	1.24	1.12	0.97	1.17	1.22	1.10	0.95	1.14	1.16	1.04	0.92	1.08	1.08	1.08	0.88	1.04
$20 000–$24 999	1.06	1.21	1.05	1.27	1.03	1.17	1.01	1.22	0.95	1.06	0.89	1.08	0.95	1.15	0.98	0.94
$15 000–$19 999	1.22	1.12	1.21	1.37	1.16	1.06	1.13	1.27	1.02	0.93	0.95	1.07	0.97	1.02	1.02	1.02
$10 000–$14 999	1.38	1.30	1.28	1.54	1.26	1.19	1.17	1.38	1.07	1.01	0.94	1.13	0.99	0.90	1.04	1.19
<$10 000	1.21	1.00	1.11	1.20	1.10	0.94	1.03	1.07	0.93	0.82	0.84	0.89	0.68	0.74	0.83	1.00

The rate ratios are from a Poisson regression model. 95% confidence intervals are available from authors on request.

Table 4 Reduction in deviance statistic for adding income to: (a) model including just age and ethnicity; b) model including age, ethnicity, marital status, education, car access, and small area socioeconomic deprivation[a]

Income Transformations	Reduction in deviance statistic			
	1981–1984	1986–1989	1991–1994	1996–1999
(a) Age and ethnicity adjusted model				
Males				
Non-transformed	85.9	86.6	273.8	264.4
Log-transformed	**94.6**	**91.9**	274.4	298.7
Rank-transformed	85.8	85.3	280.8	299.6
Females				
Non-transformed	71.2	115.9	133.8	186.0
Log-transformed	74.4	105.3	133.0	196.5
Rank-transformed	72.1	**112.0**	133.3	196.4
(b) Multivariable adjusted model[b]				
Males				
Non-transformed	39.2	30.1	93.4	75.3
Log-transformed	**45.2**	**34.0**	84.9	85.3
Rank-transformed	40.2	30.4	**93.6**	**89.7**
Females				
Non-transformed	22.6	45.7	42.6	61.5
Log-transformed	22.9	37.8	40.1	62.6
Rank-transformed	22.9	**42.9**	40.7	64.1

[a] Underlined numbers are those with at least 5% greater reduction in deviance comparing non-transformed versus log-transformed income (i.e. a test of non-linearity). **Bolded** numbers are those with at least a 5% greater reduction in deviance comparing log-transformed versus rank-transformed income. People whose household equivalized income was less than or equal to the 2.5th percentile, or greater than or equal to the 97.5th percentile were excluded from these analyses. The reduction of deviance has a Chi distribution with one degree of freedom, giving a one-sided *P*-value < 0.05 if the reduction in deviance is greater than 3.84 (i.e. all reductions are highly statistically significant).

[b] Equivalent to Model 2 in Table 3.

reduction in deviance statistic is relative to a model including just age and ethnicity (i.e. equivalent model to that presented in Table 2) and relative to a model including age, ethnicity, marital status, education, car access, and deprivation (i.e. equivalent to model 2 in Table 3 and plots in Figure 1). Comparing the reduction of deviance in the age- and ethnicity-adjusted models, it was at least 5% larger for one of the non-transformed or the log-transformed income variables in five of the eight possible comparisons (underlined numbers in first panel of Table 4). In four of these five instances (males 1981–1984, 1986–1989, and 1996–1999, females 1996–1999) the reduction in deviance was larger for the log-transformation, favouring a linear association of mortality with the logarithm of income. In the equivalent comparisons for the multivariable-adjusted model, the reduction in deviance was at least 5% larger for either the non-transformed or the log-transformed income variables in six of the eight possible comparisons, although the largest reductions were now evenly shared between the non-transformed and log-transformed income variable.

Comparing the rank-transformed and log-transformed models, the larger reductions in deviance (shown by bolded numbers in Table 4) were evenly shared across the two models and both sexes.

Discussion

Our study utilizes four cohort studies of the entire New Zealand population. The strengths of the study include its large size, and the ability to control for several socioeconomic factors. Potential weaknesses include the fact that approximately 30% of eligible mortality records were not anonymously and probabilistically linked back to a census record. However, we conducted analyses using weights[24] that adjust for potential linkage bias that may result.

Income data was collected using bin categories, prohibiting (for example) accurate determination of income among those people with a high income. Also, there was only one income question in 1986, 1991, and 1996 and it required respondents to estimate their total income from all sources combined (e.g. wage and welfare payments combined). The likely resultant mismeasurement of income presumably means that we have underestimated the magnitude of the income–mortality association. Self-employed people in New Zealand are known to declare lower incomes than might be expected on the basis of their expenditure.[28] Assuming these self-employed people also have a lower mortality risk, this systematic underestimation of their income, and perhaps other measurement errors for income, may explain the flattening of the income–mortality association at low incomes.

Confounding of the income–mortality association

The age- and ethnicity-adjusted association of income with mortality was strong. Controlling for a range of potentially confounding socioeconomic factors reduced the strength of the association, and further control for labour force status (model 3, Table 3) largely removed any income–mortality association. Our findings controlling for potential confounders were similar to those reported by Martikainen *et al*., although they did not observe such a large reduction in the income association when controlling for labour force status.[3] Labour force status is a proxy for poor health—an obvious intermediary between any socioeconomic determinant and death. Thus, controlling for labour force status may unduly bias downward the estimated income–mortality association. On the other hand, labour force status is a determinant of income, and probably has an independent association with all-cause mortality (definitely with suicide mortality[29]), and therefore is a confounder of the income–mortality association. In model 4 in Table 3 we simply restricted the analysis to employed adults as an alternative attempt to remove confounding by labour force status. However, this analysis may still be biased as exit from the labour force due to poor health is usually,[30,31] although not always,[32] reported to be more likely among people of low socioeconomic position than people of high socioeconomic position. Assuming the former situation applied in New Zealand during the 1980s and 1990s, an analysis restricted to the actively employed will underestimate the true income–mortality association.[23]

An additional reason to adjust for labour force status (as a proxy for health status) is that it will adjust for 'reverse causation' or 'health selection', a form of confounding whereby poor health leads to lower incomes giving rise to a spurious association of income with mortality. At the risk of over-generalizing, as females were less likely to be the main income earner in the household (particularly among the middle age groups that drive the mortality rates, and during the 1980s), reverse causation should be less of a bias for the female *household* income–mortality association. Inspection of rate ratios in Table 3 (models 1 and 2) and plots in Figure 1 does indicate a weaker association of household income with mortality among females, but it is not a null association. Elsewhere, we have used information on sickness benefit receipt and hospitalizations prior to census night for the 1991 NZCMS census cohort to explore the impact of health selection on the income–mortality association. (ref. 23, pp. 240–54) Whilst not definitive, these analyses also suggested that health selection may explain some, but not all, of the income–mortality association.

None of the models in Table 3 adjust for confounding in an ideal manner. We choose Model 2 as our preferred model for the following reasons. First, it is unlikely that education and marital status would be major components on the causal pathway from income to mortality. For example, education is a determinant of income, and it is usually fixed by the age of 25 so that it is unlikely to be on the causal pathway from adult income to mortality. Second, the inclusion of small area socioeconomic deprivation and car access may introduce some over-control as, for example, one's income is likely to influence where one lives and whether a person can afford to own a car. However, these two variables also probably have independent associations with mortality. Also, any over-control we have introduced is likely to be offset by our inability to control for other confounders on which we had no data. Third, adjustment for labour force status, on balance, probably over-adjusts the income–mortality association.

Arguably, a conservative interpretation of our data regarding confounding is that about half of the age- and ethnicity-adjusted association of household income with mortality was due to confounding by other socioeconomic factors, while an unknown further proportion may be due to confounding by other factors including lifestyle. However, it is very difficult to empirically identify the direct effects of any variable (such as income on mortality) in the presence of other variables that are simultaneously both confounders and mediators,[12,13] as is undoubtedly the case for most lifestyle behaviours. Furthermore, measurement error of both income and covariates may bias multivariable analyses seeking to identify independent (and hence causal) effect sizes.[13,33,34]

The shape of the income–mortality association

Logic dictates that a flattening of the association of absolute income with mortality risk must occur at some point with increasing income—mortality rates can never be less than zero, no matter how much income people earn. The plots in Figures 1a and 1b concur with this expectation, with flattening of the association of absolute income with mortality at higher incomes. Further, the association of mortality with the logarithm of income (Figures 1c and 1d) appeared to be a better linear fit—except at low incomes where the association flattens. Thus, our findings are consistent with previous work by Backlund *et al*. (1996).[1]

We also attempted to statistically test whether absolute or log-income provided the best fit with mortality using the reduction in deviance statistic from Poisson regression models (Table 4). Under this test, log-transformed tended to perform better than non-transformed income for the age- and ethnicity-adjusted models, but there was no clear 'winner' for our preferred multivariable model. However, it is critical to note that Poisson regression models (or any other standard multiplicative regression model) do not directly correspond to what we mean conceptually by a linear association of income (or any transformation of income) with mortality. We contend that hypothesized linearity of income–mortality, for any transformation of income, should be conceptualized as linearity with the absolute mortality rate (e.g. as reflected in our plotting of the rate ratios on the y-axes in Figure 1), *not* linearity with the logarithm of the mortality rate (as assumed in any multiplicative regression model.) By using a continuous specification of the three income transformation variables, the regression models are essentially equivalent to a log-scale for the rate ratios in Figure 1. In Figure 1, most of the rate ratios range from 0.75 to 1.25. Treating 1.0 as the central point, the difference between the logarithm of 1.0 and the logarithm of 0.75 is 0.2877, 29% greater than the difference between the logarithm of 1.25 and the logarithm of 1.0 of 0.2231. Thus, a good fit in these regression models used to calculate the change of deviance statistic actually corresponds to a slight upwardly concave association of the mortality rate (additive or absolute scale) with the income variable. Accordingly, a greater reduction in the deviance for the log-transformation of income (compared with non-transformed income) is strong evidence of a better linear fit for the logarithm of income. In contrast, a (modestly) greater reduction in the deviance for the non-transformed income variable is still consistent with an equivalent, or even better, linear fit for the logarithm of income.

Therefore, interpreting the change in deviance statistics accordingly, and as a supplement to the more important visual interpretation of the plots in Figure 1, we conclude that the association of the logarithm of income with mortality appears to be a better linear fit than absolute income with mortality. Our results are also consistent with the ranking of people by equivalized household being (mostly) linearly associated with all-cause mortality.

One deduction we can make is that for *both* the income rank and the logarithm of income to have a perfect linear association with mortality the distribution of people by income in society must be directly proportional to the reciprocal of income. For example, if 1000 people had an income of \$10 000, then 500 people have an income of \$20 000, 333 people have an income of \$30 000, and so on. This distribution of individuals constantly decreases as income increases. Whilst in all societies income distributions are highly skewed to the right, no income distributions have the most common income level as the lowest income level. Put another way, the modal income is always greater than the least income. Hence, *a priori* we expect that at least one of the logarithms of income or the rank-transformed income associations to change slope at low incomes. In our graphical presentation, the logarithm of income association flattened more notably at low-incomes than the income rank (Figure 1).

Martikainen *et al*. have reported that the association of absolute income with mortality in Finland is linear.[3] We propose a refinement to their conclusion. An inspection of their findings (Table 1 and Figure 1 or their paper) supports a linear association of the *decile* odds ratios with mortality, but on an absolute income scale the association flattens at higher incomes. That is, their results are actually entirely consistent with a linear association of income rank (and possibly the logarithm of income) with mortality.

Finally, we conclude that it does appear robust to use summary estimates of social inequalities in health based on income rank to compare trends over time and between populations. This is an important methodological validation of the use of the relative index of inequality (RII) which is commonly used to compare social inequalities in health, and is premised on a linear assumption of rank with health.[16]

Acknowledgements

The New Zealand census-mortality study (NZCMS) is funded by the Health Research Council of New Zealand, and co-funded by the Ministry of Health. The NZCMS is conducted in collaboration with Statistics New Zealand and within the confines of the Statistics Act 1975.

References

1 Backlund E, Sorlie PD, Johnson NJ. The shape of the relationship between income and mortality in the United States: Evidence from the National Longitudinal Mortality Study. *Ann Epidemiol* 1996;**6:**12–20.

2 Ecob R, Davey Smith G. Income and health: what is the nature of the relationship? *Soc Sci Med* 1999;**48:**693–705.

3 Martikainen P, Makela P, Koskinen S, Valkonen T. Income differences in mortality: a register-based follow-up study of three million men and women. *Int J Epidemiol* 2001;**30:**1397–405.

4 Sorlie P, Backlund E, Keller J. US mortality by economic, demographic, and social characteristics: the national longitudinal mortality study. *Am J Public Health* 1995;**85:**949–56.

5 Blakely T, Woodward A, Pearce N, Salmond C, Kiro C, Davis P. Socioeconomic factors and mortality among 25–64 year olds followed from 1991 to 1994: The New Zealand Census-Mortality Study. *NZ Med J* 2002;**115:**93–97.

6 Bucher H, Raglan D. Socioeconomic indicators and mortality from coronary heart disease and cancer: a 22-year follow-up of middle age men. *Am J Public Health* 1995;**85:**1231–36.

7 McDonough P, Duncan G, Williams D, House J. Income dynamics and adult mortality in the United States, 1972 through 1989. *Am J Public Health* 1997;**87:**1476–83.

8 Lantz P, House J, Lepkowski J, Williams D, Mero R, Chen J. Socioeconomic factors, health behaviours, and mortality. *JAMA* 1998; **279:**1703–08.

9 Mayer S. *What Money Can't Buy: Family Income and Children's Life Chances*. Cambridge, MA: Harvard University Press, 1997.

10 Robins J, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;**3:**143–55.

11 Poole C, Kaufman J. What does standard adjustment for downstream mediators tell us about social effect pathways? [Conference poster, Society for Epidemiologic Research, June 15–17 2000, Seattle]. *Am J Epidemiol* 2000;**151:**Abstract 208.

12 Cole S, Hernán M. Fallibility in estimating direct effects. *Int J Epidemiol* 2002;**31:**163–65.

13 Blakely T. Estimating direct and indirect effects: fallible in theory, but in the real world? *Int J Epidemiol* 2002;**31:**166–67.

14 Gravelle H. How much of the relation between population mortality and unequal distribution of income is a statistical artefact? *BMJ* 1998; **316:**382–85.

15 Wagstaff A, van Doorslaer E. Income inequality and health: What does the literature tell us? *Annu Rev Public Health* 2000;**21:**543–67.

16 Mackenbach J, Kunst A. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Soc Sci Med* 1997;**44:**757–71.

17 Kunst A, Groenhof F, Mackenbach J, The EU working group on socioeconomic inequalities in health. Mortality by occupational class among men 30–64 years in 11 European countries. *Soc Sci Med* 1998; **46:**1459–76.

18 Kunst A, Groenhof F, Mackenbach J, the EU Working Group on Socioeconomic Inequalities in Health. Occupational class and cause specific mortality in middle-aged men in 11 European countries: comparison of population based studies. *BMJ* 1998;**316:**1636–42.

19 Blakely T. *Socio-economic Factors and Mortality among 25–64 Year Olds: The New Zealand Census-Mortality Study*. (Also at http://www.wnmeds.ac.nz/nzcms-info.html) [Doctorate]. University of Otago, 2001.

20 Wilkinson R. *Unhealthy Societies. The Afflictions of Inequality*. London: Routledge, 1996.

21 Blakely T, Salmond C, Woodward A. *Anonymous Record Linkage of 1991 Census Records and 1991–94 Mortality Records: The New Zealand Census-Mortality Study* (Also at http://www.wnmeds.ac.nz/nzcms-info.html). Wellington: Department of Public Health, Wellington School of Medicine, University of Otago, 1999.

22 Blakely T, Salmond C, Woodward A. Anonymous linkage of New Zealand mortality and Census data. *Aust NZ J Public Health* 2000; **24:**92–95.

23 Blakely T. *The New Zealand Census-Mortality Study: Socioeconomic Inequalities and Adult Mortality 1991–94. Also at* http://www.wnmeds.ac.nz/nzcms-info.html. Wellington: Ministry of Health, 2002:258.

24 Fawcett J, Blakely T, Atkinson J. Weighting the 81, 86, 91 & 96 census-mortality cohorts to adjust for linkage bias. *NZCMS Technical Report No. 5*. ISBN 0–473–09112–7 (Also at http://www.wnmeds.ac.nz/nzcms-info.html). Wellington: Department of Public Health, Wellington School of Medicine and Health Sciences, University of Otago, 2002:92.

25 Jensen J. *Income Equivalences and the Estimation of Family Expenditure on Children*. Wellington: Department of Social Welfare (unpublished), 1988.

26 Salmond C, Crampton P, Sutton F. NZDep91: A New Zealand index of deprivation. *Aust NZ J Public Health* 1998;**22:**835–37.

27 Salmond C, Crampton P. NZDep96—what does it measure? *Social Policy Journal of New Zealand* 2001;**17:**82–100.

28 Davis P, Jenkin G, Coope P, Blakely T, Sporle A, Kiro C. The New Zealand Socioeconomic Index of Occupational Status: methodological

revision and imputation for missing data. *Aust NZ J Public Health* (in press).

29 Blakely T, Collings S, Atkinson J. Unemployment and suicide: New Zealand evidence for a causal association. *J Epidemiol Community Health* 2003;**57:**594–600.

30 Power C, Matthews S, Manor O. Inequalities in self rated health in the 1958 birth cohort: lifetime social circumstances or social mobility. *BMJ* 1996;**313:**449–53.

31 van de Mheen H, Stronks K, Schrijvers C, Mackenbach J. The influence of adult ill health on occupational class mobility and mobility out of and into employment in The Netherlands. *Soc Sci Med* 1999;**49:**509–18.

32 McDonough P, Amick BC 3rd. The social context of health selection: a longitudinal study of health and employment. *Soc Sci Med* 2001;**53:**135–45.

33 Phillips A, Davey Smith G. How independent are 'independent' effects? Relative risk estimation when correlated exposures are measured imprecisely. *J Clin Epidemiol* 1991;**44:**1223–31.

34 Davey Smith G, Phillips A. Declaring independence: why we should be cautious. *J Epidemiol Community Health* 1990;**44:**257–58.