The New Zealand Socioeconomic Index: developing and validating an occupationally-derived indicator of socio-economic status

Peter Davis

Department of Community Health, School of Medicine, University of Auckland, New Zealand

Keith McLeod

Statistics New Zealand, Wellington

Miriam Ransom

Department of Community Health, School of Medicine, University of Auckland, New Zealand

Patrick Ongley

Statistics New Zealand, Wellington

Neil Pearce

Department of Medicine, Wellington School of Medicine, University of Otago, New Zealand

Philippa Howden-Chapman

Department of Public Health, Wellington School of Medicine, University of Otago, New Zealand

Abstract

Objective: Following revision of the international standard classification (ISCO88), to update and validate on health data an occupationally derived indicator of socio-economic status (SES) adapted to changing occupational and demographic conditions.

Method: The development of the New Zealand Socioeconomic Index (NZSEI) is based on a 'returns to human capital' model of the stratification process and uses data from the 1991 New Zealand Census (n=1,051,926) to generate scores for 97 occupational groups. The construct validation of the scale is carried out on data from the 1992-93 nationwide Household Health Survey (n=3,000) using three health indicators (self-assessed health, cigarette smoking, general practitioner visits).

Results: In general, the results are consistent with expected socio-economic patterns drawn from the literature for the three indicators.

Conclusions: While further work is required on a number of methodological and conceptual issues, the NZSEI provides a robust, standardised and internationally comparable occupational scale of SES for both males and females in either full- or part-time employment.

Implications: The NZSEI can be used on routinely collected occupational data. It has a clear conceptual rationale, updates existing SES scales, and provides a link to international standards in SES and occupational classification.

(*Aust N Z J Public Health* 1999; 23: 27-33)

Occupational measures of socio-economic status (SES) are routinely used in public health research. Indeed, aside from gender and age, occupational SES is probably the most widely invoked explanatory variable in such research.[1] Furthermore, occupational information – such as that contained in death certificates – is frequently the only socio-economic data available.[2]

To date, the occupational SES measure used most widely within the New Zealand research community has been the Elley-Irving (E-I) scale.[3-5] However, given the uncertain conceptual rationale for that scale, and following the shifts which have occurred over the past decade in the occupational structure, as well as broader demographic and social changes, there appeared to be good grounds for developing a new local instrument. This paper presents the results of such an exercise in the development and construct validation of the New Zealand Socioeconomic Index (NZSEI), an occupational scale of SES.

The rationale of this investigation rests on two premises. First, it is assumed that a person's occupation is a reasonable basis on which to allocate them in the socio-economic hierarchy.[6] This is the conceptual basis for the *development* of the scale. Second, the assumption is made that differences in life chances and lifestyles flowing from placement in the socio-economic hierarchy will be reflected in the patterning of key health indicators.[7] This is the analytical rationale behind the *validation* of the NZSEI.

The *development* of the NZSEI is premised on a 'returns to human capital' model of the stratification process.[8,9] Specifically, it is proposed that there exists a fundamental relationship between cultural capital or resources (education) and access to material rewards (income), and that this relationship is mediated through the occupational structure (see Figure 1).

The *validation* of the NZSEI relies on the concept of construct validity;[10] that is, testing whether relationships predicted by theory are empirically borne out – in this case,

Correspondence to:
Dr Peter Davis, Department of Community Health, School of Medicine, University of Auckland, Private Bag 92019, Auckland, New Zealand. Fax: +64 9 373 7503; e-mail: pb.davis@auckland.ac.nz

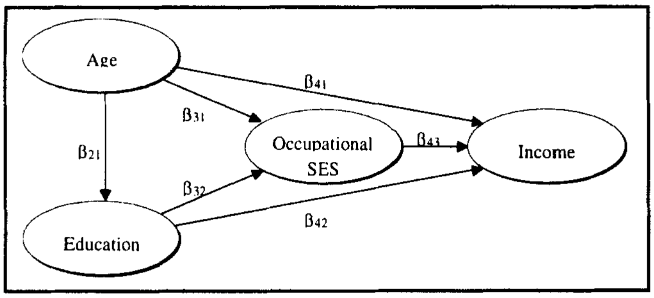


Figure 1: 'Returns on Human Capital' Path Model of Occupational Stratification.

whether known and theoretically informed patterns of socio-economic determination for a range of health indicators are confirmed.[11] Drawing on data from the 1992-93 Household Health Survey (HHS), three measures are used – in health status, health-related behaviour and lifestyles, and receipt of health services, respectively. The reliability of the NZSEI will also be assessed by comparing the results with those obtained from the E-I scale.

Methods

The Statistical Algorithm

Putting the 'returns on human capital' concept of the stratification process into operation involves specifying a path model represented by a series of regression equations and generating scores for the unmeasured variable, occupation. These equations are estimated iteratively in such a way as to minimise the regression coefficient of income on education – the direct effect – and maximise the indirect effect of the same relationship mediated through occupation, generating scores in the process.[12] Because of its known confounding effect, age is also included in the model.

The model is represented in diagrammatic form in Figure 1. The arrows linking the variables are represented in the algorithm as regression coefficients, and the model as a whole is represented by a series of linear regression equations. The three explanatory variables age, income, and education, are used to derive a value for the unknown occupational SES variable. This is calculated in such a way as to minimise the regression coefficient (β_{42}) directly linking income with education. This is not worked out exactly, or in one step, but uses an iterative optimal scaling technique known as an alternating least squares algorithm to approximate scores which result in a minimal β_{42}. The result is a series of scores for each occupation which represent an optimally weighted combination of income and education variables, corrected for age, based on the assumptions of our model (i.e. no direct education-income link). These are the values of the NZSEI; they represent the scores for an occupational SES scale.

The alternating least squares algorithm is summarised in Figure 2. The age, income, and education variables are represented by a, i and e. Occupational SES scores are represented by o, where o' and o" are revisions of this within each iteration of the algorithm, and $g_1 \ldots g_k$ are the SES scores for the k occupations. The SES scores are finally scaled on a continuum from 10 to 90. The algorithm for deriving the NZSEI scores is outlined in more detail in the Appendix.

Step 1:	Initialise the education and income weights at any reasonable starting point, and from these construct an initial o.	
Step 2:	Regress i on a and o → Regress o on e and a → Regress e on a →	β_{41} , β_{43} β_{31} , β_{32} β_{21}
Step 3:	Compute $o' = \beta_{43}(i - \beta_{41}a) + \beta_{32}e + \beta_{31}a$ Standardise o' Compute scores as means of o' for $\gamma_1, \ldots \gamma_k$ Compute o" using the new scaling	
Step 4:	Regress i on a, e and o" → If minimum on β_{42} , step out Go back to step 2 and substitute o" for o	β_{42}

Figure 2: The NZSEI Algorithm.

Dataset – Scale Development

The NZSEI was developed using data from the 1991 Census of Population and Dwellings. The census database contains the individual records of 3,373,926 resident New Zealanders. Age, income and educational information was drawn from this database for male and females in the full-time workforce over the age range 21 to 69. This reduced the dataset to 1,051,419 persons.

For those currently in the workforce, occupational data in the Census were coded into the New Zealand Standard Classification of Occupations 1990 (NZSCO90).[13] This is based on the International Standard Classification of Occupations 1988 (ISCO88) and replaces the previous New Zealand standard, NZSCO68.[14] The occupational data for the 1991 Census was also coded into NZSCO68, the occupational standard for the E-I scale, thus permitting conversion of scores between NZSEI and E-I.

The NZSCO90 classification contains 10 major groups, which are sub-divided into 24 sub-major groups, 97 minor groups, 260 unit groups and 563 groups. The NZSEI is based on aggregations at the minor group (or three-digit) level because the breakdown of occupations at this level is detailed enough to be of utility in social research and also based on sufficient numbers within each occupational category to give stable estimates. E-I scores were allocated according to the 1981 Census version.[5]

The income variable in the Census dataset records total personal income for the year preceding the Census and is collected within bands. For the purposes of the analysis a mid-point was taken in each of these bands and this value was assigned as an income score to all individuals in that band. In the open-ended top income bracket a mid-point was assigned using data from the Household Economic Survey, which records exact income values. In order to normalise the skewed distribution of income scores, logs were taken and the resulting values used in the algorithm.

The education variable in the 1991 census dataset was represented by qualifications achieved. In order to obtain a common metric for the purposes of the statistical analysis, these were converted into equivalent years of formal education ranging from 10 years – for those with no qualifications – through to 19 years (for postgraduate qualifications).

Dataset – Construct Validation

The data used for the construct validation exercise were from Statistics New Zealand's 1992-93 HHS.[15] More than 7,000 New Zealanders were surveyed during a 12-month period from the usually resident, non-institutionalised, civilian population living in private households. We further narrowed the scope for our analyses to the full-time workforce, resulting in a dataset of almost 3,000.

Three variables were derived using the following questions:

(a) *Self-assessed health*: "Overall would you say your health is excellent, good, not so good, or poor?".

(b) *Smoking*: "Would you describe yourself as a tobacco smoker, an ex-smoker, or a non-smoker?"

(c) *GP visits*: Since this time last year, how many times have you seen a doctor or been visited by a doctor?".

For the purposes of the analysis these items were converted into binary variables and percentages as follows: (a) per cent "poor" or "not so good" health; (b) per cent who were currently smokers; (c) percent who had visited a general practitioner (GP) in the previous year.

Since the HHS dataset only contained NZSCO90 occupation codes, E-I scores could not be assigned directly. Many NZSCO90 codes spanned more than one E-I class. In these cases the E-I assignment was based on a proportionate allocation, with respondent income as recorded in the HHS being the final determining factor.

Table 1: NZSEI Scores for occupational groups: Sub-Major (2-digit) Level, NZSCO90.

Code	Description	NZSEI
01	Armed Forces	54
11	Legislators and Administrators	77
12	Corporate Managers	57
21	Physical, Mathematical and Engineering Science Professionals	74
22	Life Science and Health Professionals	71
23	Teaching Professionals	69
24	Other Professionals	71
31	Physical Science and Engineering Associate Professionals	62
32	Life Science and Health Associate Professionals	55
33	Other Associate Professionals	56
41	Office Clerks	43
42	Customer Services Clerks	41
51	Personal and Protective Services Workers	38
52	Salespersons, Demonstrators and Models	33
61	Market Oriented Agricultural and Fishery Workers	25
71	Building Trades Workers	47
72	Metal and Machinery Trades Workers	49
73	Precision Trades Workers	48
74	Other Craft and Related Trades Workers	38
81	Industrial Plant Operators	48
82	Stationary Machine Operators and Assemblers	34
83	Drivers and Mobile Machinery Operators	40
84	Building and Related Workers	44
91	Labourers and Related Elementary Service Workers	31

Results

Scale Development

The NZSEI was derived at the minor group (3-digit) level of the 1990 version of the New Zealand Standard Classification of Occupations (NZSCO90) which comprises 97 occupations. The results were scaled on a 10 to 90 range for consistency with the international index. The three-digit results were used to derive sub-major group (two-digit) and major group (one-digit) results; at these levels there are 24 and 10 occupational classes respectively. To get NZSEI values at the higher levels of aggregation (major and sub-major groups), a weighted mean of the constituent three-digit occupation NZSEI scores was calculated. Table 1 shows the index scores at the sub-major group (2-digit) level. Results at the minor group (3-digit) level are available from the authors.

There are advantages and disadvantages to a continuous scale, and for many applications a discrete scale is preferable. We split the NZSEI into 6 groups, firstly as a starting point for users who prefer discrete classes, and secondly to provide a basis both for the NZSEI construct validation work and for the comparisons with the E-I scale (see Figure 4 and Table 3).

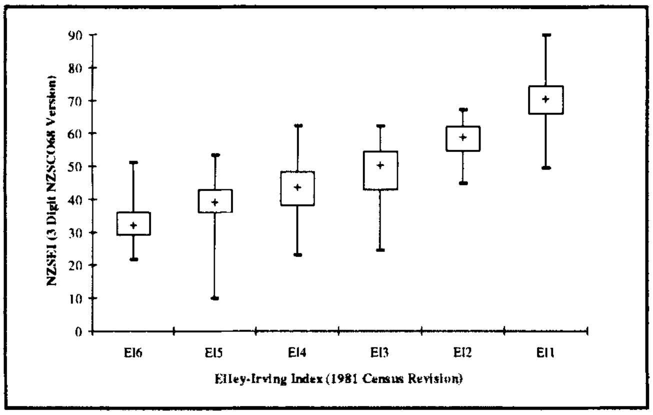
The groups were split in such a way as to ensure each class consisted of a reasonable proportion of the population. Apart from this consideration, the splits were done at a point which was fairly arbitrarily assigned. The wider range of 20 was assigned to class 6 as there was only one small occupation group (Non-Ordained Religious Associate Professionals) with an NZSEI score of less than 20. The breakdown of the workforce into each of the six NZSEI 'classes' is presented in Table 2.

Since the NZSEI was produced in part as an update for the E-I scale, we wanted to look at the relationship between the two measures. As the E-I Index was created using NZSCO68, comparisons between the two scales were carried out using the NZSEI scores based on the NZSCO68 occupation groups. Figure 3 shows the distribution of NZSEI scores for people who were assigned each E-I category.

The box and whisker plot shows the median NZSEI value as a cross, inside a box whose top and bottom lines represent the upper and lower quartiles of NZSEI values respectively. The lines

Table 2: Distribution of full-time workforce over NZSEI Classes.

Class	NZSEI Range	% of Population
1	75-90	5.0
2	60-74	17.5
3	50-59	20.7
4	40-49	22.8
5	30-39	16.5
6	10-29	17.5



ABOVE: Figure 3: Box Plots: Distribution of NZSEI Scores by Elley-Irving Class.

BELOW: Figure 4: Distribution of health indicators by SES Class (NZSEI and E-I).

projecting from the top and bottom of the box end at the maximum and minimum NZSEI points. The figure shows that, while there is a broad consistency between NZSEI and E-I scores, there is some overlap in the ranges, especially for the middle SES groups.

Construct Validation

As a comparison of how the new scale compares to the E-I Index – and as a form of construct validation – we produced dot charts for each health variable against both NZSEI and E-I (see Figure 4).

The upper panel of Figure 4 shows smoking levels plotted for each of the six E-I and NZSEI classes. There is a very similar distribution for both scales; a strong and more or less consistent gradient, except for the reversal at the lower end of the distribution. Both scales are again closely similar for self-assessed health

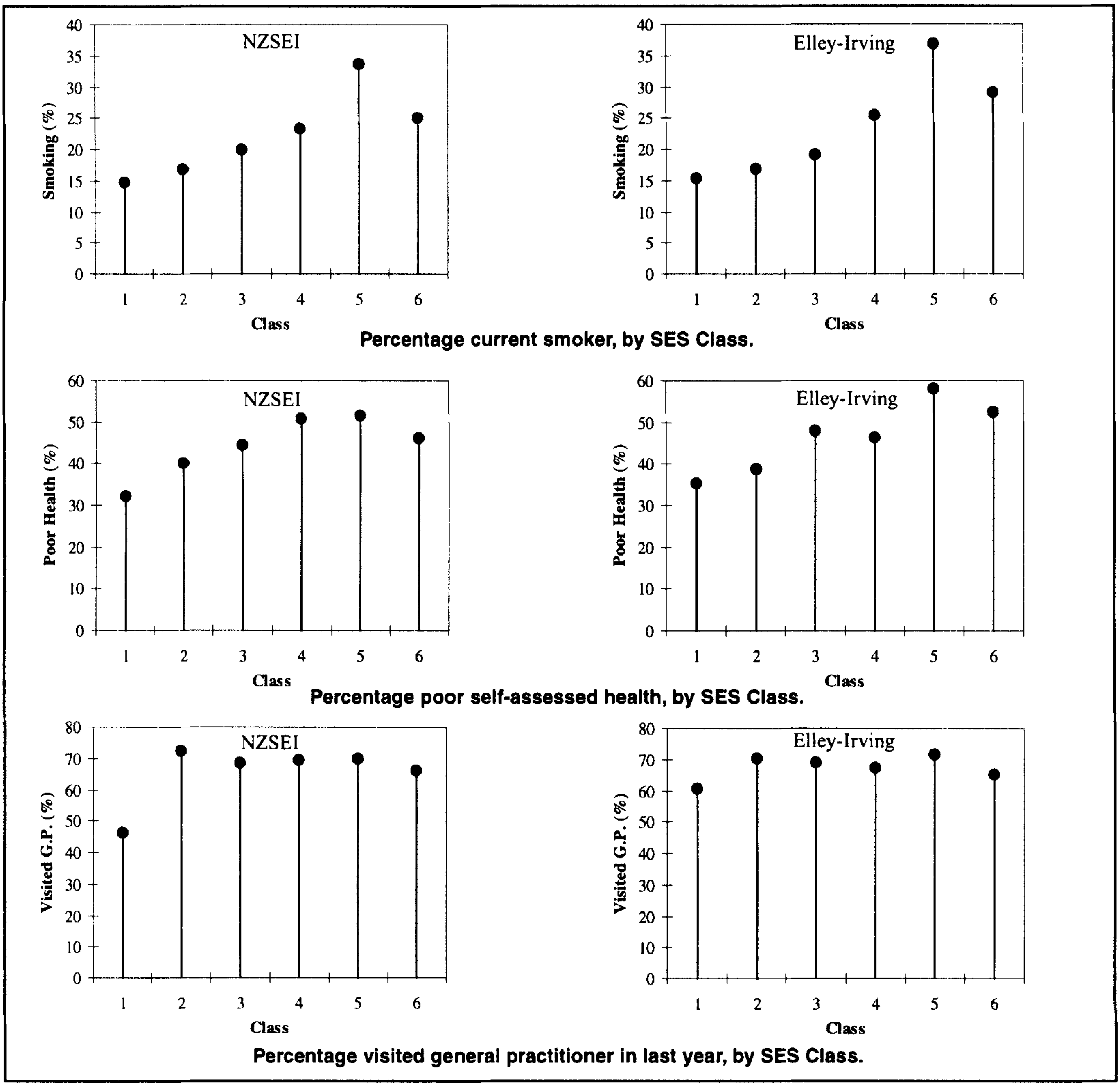


Table 3: Summary measures of effect for three health indicators, by NZSEI and E-I.

	Current Smoker		Poor Health		GP Visits	
Measure	**NZSEI**	**E-I**	**NZSEI**	**E-I**	**NZSEI**	**E-I**
Rate Ratio[a]	1.70	2.13	1.43	1.65	1.30	1.09
Rate Ratio[b]	1.76	2.14	1.27	1.53	1.06	1.03
Rate Difference[a]	10.3	17.3	14.0	22.7	15.4	5.64
Rate Difference[b]	12.5	18.7	10.3	20.1	3.73	1.70
Relative PAR (%)	37.7	36.0	30.5	26.1	24.7	11.9
Absolute PAR	8.89	8.62	14.2	12.4	16.8	8.15

Notes:
(a) Measures the difference between the top group and the bottom group.
(b) Measures the difference between the top two groups and the bottom two groups.

in the middle panel of the figure, although the gradient is not nearly as marked. Finally, the pattern of similarity continues in the lower panel, with a much weaker gradient; group one has a lower rate of general practitioner visiting in the previous year, but thereafter the rate seems to be relatively uniform.

Some simple summary measures of the magnitude of inequalities across this range of socio-economic comparisons are presented in Table 3. The rate ratio measures the ratio of the rate of the lowest SES group to that of the highest group. A second rate ratio is presented, showing the comparison between the top two and bottom two groups. The rate difference, a measure of the absolute difference between the rates of the top and bottom groups, is similarly shown for these two splits. Finally, population-attributable risk (PAR) measures were produced. These represent the proportional or absolute reduction in rates that would occur in the hypothetical case that everyone had the rates of the highest SES group.

As was apparent from the dot charts in Figure 3, the measures indicate that the E-I scale exerts greater discriminatory power over the range of the classes, except possibly for general practitioner visits. All of the measures show a fairly significant socio-economic effect, although as expected from Figure 3, general practitioner visits does so only for NZSEI and for comparisons between the top and bottom single groups.

Discussion

This paper has developed, and provided a construct validation for, an occupationally-derived socio-economic index that provides a number of advances over the existing, and widely used, E-I scale. While controversy remains about the best way to measures social class,[16] the NZSEI has a number of strengths: it is up-to-date, using data from the recent 1991 Census and the NZSCO90; it goes beyond the male full-time workforce to include women and thus derives scores of equal applicability to both males and females;[17] its results are applicable to part-timers; the more rigorous conceptual basis of the scale facilitates a greater theoretical understanding of the mechanisms linking SES and social outcomes such as health status;[18] a considerably more advanced and generic statistical model has been applied in order to derive the weights and scores that form the basis of the scale;[8] finally, there are direct linkages to an international equivalent – ISEI – and to the international standard occupational coding system, ISCO88.[9]

The results of the NZSEI analysis were largely as expected. Income was assigned a higher weighting than education, and this was the prime SES indicator in all analyses.[19] This result is in contrast to the recent work of Hauser and Warren which gives predominant weight to education.[20] However, it should be noted that the criterion variable in this latter analysis was occupational prestige, and such rankings may be more heavily influenced by popular perceptions of the educational rather than the income attributes of occupations.

There were one or two unusual results, although these could be traced back to the income or education values in the Census dataset. Some occupations, such as religious professionals and farmers had very low income, and a correspondingly low NZSEI score. Problems of this kind appeared to occur most often with occupations with high numbers of self-employed, and this may indicate that for this group self-reported income is a somewhat unreliable indicator of material position.[21]

Other results produced on the full Census dataset of all full-time workers, both male and female, and reported elsewhere showed a reasonable 'fit' to the scores generated in the international scale.[17] Removing female workers from the New Zealand dataset – in order to approximate the international analysis more closely – improved this 'fit' considerably. Although there were systematic differences in the expected direction in occupational scores between males and females, the underlying structure of the statistical model remained the same for both groups. A similar result held for the analysis by ethnic group. Adding part-time workers made little difference to the results. Although the final scores relate to the full-time workforce – males and females, all ethnic groups – they can also be accepted as closely indicative for part-timer workers as well.

An exercise was also conducted to assess the correspondence of the NZSEI scores with those derived from the existing E-I scale. Although there was a clear association between the two scales, there was also a wide spread of NZSEI scores for each E-I group. Differences between the scales could be due to the different models used, changes in the population between the time the scales were derived, or due to a change in population specifications (the E-I scale was based on data for males only, and also was restricted to those aged 25 to 44 years old).[3-5]

Central to the model of the stratification process – and underpinning the construct validation – is the assumption that inequalities in the occupational order translate into corresponding variations in health chances and health outcomes. In order to carry out this analysis a categorical version of the NZSEI scale was used with occupations allocated to six groups. Except for the lowest group – which departs from the expected pattern – both smoking and self-assessed health demonstrate class gradients, as expected from the literature.[22,23] In the case of general practitioner visits, however, the relationship was much less clear-cut (which also has support in the literature).[24]

A further set of comparisons were carried out with the E-I scale. Comparisons with E-I using simple dot charts showed that NZSEI replicated the same patterns of relationship between SES and health outcome, although perhaps with less discriminatory power.

Conclusion

This paper has developed and tested a socio-economic index using an established statistical methodology. Although tested on certain health outcomes, it has been developed with a view to its widest possible application in social science research and official statistics. These scores have been generated on the most complete dataset available and detailed information has been provided for researchers wishing to use the scale. Linking data has been provided to earlier statistical series and to the existing E-I scale. Further work is clearly required on a number of issues, including rural occupations, the place of those outside the workforce, the position of women, household versus individual measures, and a categorical scale. In the meantime, however, the NZSEI, a continuous occupational scale of SES, provides a robust, standardised and internationally comparable measure of occupational class.

Acknowledgments

This project was funded by the Health Research Council of New Zealand. We are grateful to Len Cook, Government Statistician, and Denise Brown, Chief Analyst Social Policy, for facilitating the development and progress of this project. Other officers of Statistics New Zealand involved with this project have been Max Wigbout, Minoo Meimand and Jenny Mason. The views expressed in this paper are those of the authors and do not necessarily represent those of Statistics New Zealand.

References

1. Marmot M, Kogevinas M, Elston M. Social/economic status and disease. *Am Rev Public Health* 1987; 8: 111-35.
2. Najman J, Bampton M. An ASCO-based occupational status hierarchy for Australia: a research note. *Aust N Z J Soc* 1991; 27: 218-31.
3. Elley W, Irving J. A socio-economic index for New Zealand based on levels of education and income from the 1966 census. *N Z J Educ Stud* 1972; 2: 153-67.
4. Elley W, Irving J. Revised socioeconomic index for New Zealand. *N Z J Educ Stud* 1976; 11: 25-36.
5. Elley W, Irving J. The Elley-Irving socio-economic index 1981 census revision. *N Z J Educ Stud* 1985; 20: 115-28.
6. Rose D, O'Reilly K. Constructing classes. In: Rose D, O'Reilly K, editors. *Constructing classes. Towards a new social classification for the UK.* Swindon (UK): ESRC/ONS, 1997: 1-20.
7. Kunst A, Mackenbach J. *Measuring socioeconomic inequalities in health.* Copenhagen: World Health Organisation Regional Office, 1994.
8. Ganzeboom H, De Graaf P, Treiman D, de Leeuw J. A standard international socioeconomic index of occupational status. *Soc Sci Res* 1992; 21: 1-56.
9. Ganzeboom H, Treiman D. Internationally comparable measures of occupational status for the 1988 ISCO. *Soc Sci Res* 1996; 25: 201-39.
10. Aday LA. *Designing and conducting health surveys.* 2nd edition. San Francisco: Jossey-Bass, 1996.
11. Najman JM. Class inequalities in health and lifestyle. In: Waddell C, Petersen AR, editors. *Just health. Inequality in illness, care and prevention.* Melbourne: Churchill Livingstone, 1994: 27-46.
12. de Leeuw J. Multivariate analysis with linearization of regressions. *Psychometrika* 1989; 53: 437-54.
13. Department of Statistics. *New Zealand Standard Classification of Occupations 1990.* Wellington: Department of Statistics, 1992.
14. Department of Statistics. *New Zealand Standard Classification of Occupations Manual.* Wellington: Department of Statistics, 1976.
15. Statistics New Zealand/Ministry of Health. *A picture of health.* Wellington: SNZ/MOH, 1993.
16. Turrell G, Western JS, Najman JM. The measurement of social class in health research: problems and prospects. In: Waddell C, Petersen AR, editors. *Just health. Inequality in illness, care and prevention.* Melbourne: Churchill Livingstone, 1994: 87-104.
17. Davis P, McLeod K, Ransom R, Ongley P. *The New Zealand socioeconomic index of occupational status (NZSEI).* Wellington: Statistics New Zealand, 1997. SNZ Research Report No.: 2.
18. Arber S. Insights about the non-employed class and health: Evidence from the General Household Survey. In: Rose D, O'Reilly K,editors. *Constructing classes. Towards a new social classification for the UK.* Swindon (UK): ESRC/ONS, 1997: 78-92.
19. Elias P, McKnight A. Earnings, occupations and social classification. In: Rose D, O'Reilly K, editors. *Constructing classes. Towards a new social classification for the UK.* Swindon (UK): ESRC/ONS, 1997: 98-115.
20. Hauser RM, Warren JR. Socioeconomic indexes of occupational status: A review, update, and critique. In: Raftery A, editors. *Sociological methodology 1997.* Cambridge: Blackwell Publishers, 1997: 177-298.
21. Eardley T, Bradbury B. *Self-employment and social security.* Sydney: SPRC, 1997. SPRC Reports and Proceedings No.: 133.
22. Macintyre S. The patterning of health by social position in contemporary Britain: Directions for sociological research. *Soc Sci Med* 1986; 23: 393-415.
23. Blaxter M. A comparison of measures of inequality in morbidity. In: Fox J, editors. *Health inequalities in European countries.* Aldershot (UK): Gower, 1989: 199-230.
24. Macintyre S. The role of health services in relation to inequalities in health in Europe. In: Fox J, editors. *Health inequalities in European countries.* Aldershot (UK): Gower, 1989: 317-32.

Appendix – Description of the algorithm

The path model can be represented by three linear regression equations.

$$
\begin{aligned}
i &= \beta_{41} a + \beta_{42} e + \beta_{43} o + \varepsilon \\
o &= \beta_{31} a + \beta_{32} e + \varepsilon \\
e &= \beta_{21} a + \varepsilon
\end{aligned}
$$

i, e and a are normalised income, education and age variables, and o is our unknown occupational SES variable, also normalised. The beta coefficients represent the arrows on the path diagram.

By changing the values of o we alter the relationships between the three observable variables. What we want to do therefore is assign each individual an o score such that we get a minimal β_{42}. Since we are interested in an occupation-based SES scale, we have an added restriction that everyone in an occupation will have the same o score.

In order to approximate a minimal β_{42}, the loss function we minimise is the total residual sum of squares, σ_N, for the model with the education-income path represented by β_{42} left out. Each of the variables can be described as a (jx1) vector containing values for each of the j people in the population. As input to the algorithm at each iteration we have vectors $\underline{i}$, $\underline{a}$, $\underline{e}$ and the SES estimate, $\underline{o}$.

$$
\begin{aligned}
\sigma_N = &\| \underline{i} - (\beta_{41}\, \underline{a} + \beta_{43}\, \underline{o}) \|^2 \\
&+ \| \underline{o} - (\beta_{31}\, \underline{a} + \beta_{32}\, \underline{e}) \|^2 \\
&+ \| \underline{e} - \beta_{21}\, \underline{a} \|^2
\end{aligned}
$$

This can also be written as:

$$
\begin{aligned}
\sigma_N &= \sum_j [(i_j - \beta_{41}a_j - \beta_{43}o_j)^2 + (o_j - \beta_{31}a_j - \beta_{32}e_j)^2 + (e_j - \beta_{21}a_j)^2] \\
&= \sum_j [(i_j - \beta_{41}a_j - \beta_{43}h_{jk}\gamma_k)^2 + (h_{jk}\gamma_k - \beta_{31}a_j - \beta_{32}e_j)^2 + (e_j - \beta_{21}a_j)^2]
\end{aligned}
$$

The alternating least squares algorithm alternates between two steps in each iteration. The first step is to derive optimal transformations (beta coefficients) based on given SES scores. In the second step we derive SES values, optimal on our loss function, for the transformations calculated in the first step. The steps are alternated until the algorithm converges. Since in each iteration we are improving on the results of the previous iteration, the loss function we are minimising necessarily reduces. The loss function, σ_N, has a lower bound of zero, and therefore the algorithm will always converge.

To initialise the $\underline{o}^{(0)}$ vector before the first iteration of the algorithm we decide on initial income and education weights (we used 0.5 for both, but experimented with other values), and from these construct a vector of weighted averages of income and education. This is then averaged over each occupation and normalised to give us our initial SES estimate, $\underline{o}^{(0)}$.

We represent $\underline{o}^{(0)}$, which contains the occupation SES estimates for each individual in the sample, as being equal to $H\underline{\gamma}^{(0)}$, where $\underline{\gamma}^{(0)}$ is the (kx1) vector containing the SES estimates for each of the k occupations represented in the sample and H is the dummy corresponding to $\underline{o}^{(0)}$.

i.e.

$$
\begin{bmatrix} o_1 \\ o_2 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ o_j \end{bmatrix}
=
\begin{bmatrix} \gamma_1 \\ \gamma_1 \\ \vdots \\ \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \\ \vdots \\ \gamma_k \end{bmatrix}
=
\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}
\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix}
$$

After initialising $\underline{o}^{(0)}$, the next step is to calculate the beta coefficients for the three regressions, with the β_{42} path excluded. These five coefficients represent our initial beta estimates, $\beta^{(0)}$.

Next, we revise our SES estimates by minimising σ_N over $\underline{\gamma}^{(0)}$. A minimum point, $\underline{\gamma}'$ is calculated by differentiating σ_N by $\underline{\gamma}^{(0)}$.

$$
\begin{aligned}
\frac{d\sigma_N}{d\underline{\gamma}} &= \sum_j [2\beta_{43}^2 \underline{h}_j^2 \underline{\gamma} + 2\beta_{41}\beta_{43} a_j \underline{h}_j - 2\beta_{43} i_j \underline{h}_j + 2\underline{h}_j^2 \underline{\gamma} - 2\beta_{31} a_j \underline{h}_j - 2\beta_{32} e_j \underline{h}_j] \\
&= 2\sum_j [(\beta_{43}^2 + 1)\underline{h}_j^2 \underline{\gamma} - \beta_{43}(i_j - \beta_{41}a_j)\underline{h}_j - \beta_{31}a_j\underline{h}_j - \beta_{32}e_j\underline{h}_j]
\end{aligned}
$$

The resulting equation is set equal to zero, allowing us to calculate an optimal $\underline{\gamma}'$.

$$
\begin{aligned}
\underline{0} &= 2[(\beta_{43}^2 + 1)H^T H\underline{\gamma}' - \beta_{43}H^T(\underline{i} - \beta_{41}\underline{a}) - \beta_{31}H^T\underline{a} - \beta_{32}H^T\underline{e}] \\
&= (\beta_{43}^2 + 1)H^T H\underline{\gamma}' - H^T(\beta_{43}(\underline{i} - \beta_{41}\underline{a}) + \beta_{31}\underline{a} + \beta_{32}\underline{e}) \\
(\beta_{43}^2 + 1)H^T H\underline{\gamma}' &= H^T(\beta_{43}(\underline{i} - \beta_{41}\underline{a}) + \beta_{31}\underline{a} + \beta_{32}\underline{e}) \\
\underline{\gamma}' &= \frac{(H^T H)^{-1} H^T(\beta_{43}(\underline{i} - \beta_{41}\underline{a}) + \beta_{31}\underline{a} + \beta_{32}\underline{e})}{(\beta_{43}^2 + 1)}
\end{aligned}
$$

The $\underline{\gamma}'$ scores are then used to derive a vector $\underline{o}'$, which is normalised to obtain a new SES estimate, $o^{(1)}$.

$$
\begin{aligned}
\underline{o}' &= H\underline{\gamma}' \\
&= \frac{1}{(\beta_{43}^2+1)} H(H^T H)^{-1} H^T \; (\beta_{43}(\underline{i} - \beta_{41}\underline{a}) + \beta_{31}\underline{a} + \beta_{32}\underline{e})
\end{aligned}
$$

The $\beta_{42}^{(1)}$ coefficient is next calculated by regressing i on a, e, and $o^{(1)}$. The $o^{(1)}$ scores become the estimates for the next iteration, $\underline{o}^{(0)}$, and the process is repeated. The algorithm converges to a stationary β_{42}, stopped in our case when the difference between the new β_{42} estimate and the last is less than 0.01.

The technique described doesn't actually minimise β_{42}, or correspondingly maximise $\beta_{43}\beta_{42}$. Instead we minimise the residual sum of squares of the saturated model, with β_{42} included, plus the square of our usual least squares estimate for β_{42}. So, we are minimising σ_n where:

$$
\sigma_n = \min_{\beta_{42}} \sigma_s + \hat{\beta}_{42}^2
$$

While this is roughly the same as minimising β_{42}, it is not identical.

Although assigning a different criterion would result in a different solution, any such differences should not be large.