

**EXPLORANDO LOS FACTORES QUE PUEDEN EXPLICAR LA PREFERENCIA POR
LAS POLÍTICAS IGUALITARIAS EN SALUD: UN ESTUDIO EMPÍRICO**

Ignacio Abásolo ¹ y Aki Tsuchiya ²

¹ Departamento de Economía de las Instituciones, Estadística Económica y Econometría, Facultad de Ciencias Económicas y Empresariales. Universidad de La Laguna, Campus de Guajara. La Laguna. Tenerife. Fax: 0034-922317204; teléfono: 0034-922317088; e-mail: iabasolo@ull.es

² Sheffield Health Economics Group, University of Sheffield. Reino Unido.

**TRABAJO EN CURSO
(POR FAVOR, NO CITAR SIN EL PERMISO DE LOS AUTORES)**

Objetivo: Estudiar empíricamente qué factores están asociados a una actitud igualitaria frente a las desigualdades socioeconómicas en salud.

Método: Se llevó a cabo una encuesta sobre una muestra representativa de la población española (n=1209). Después de informar que los individuos de las clases sociales altas tienen una esperanza de vida al nacer mayor que los individuos de clases sociales bajas, se pide a los entrevistados que elijan entre dos políticas de salud: una política que aumentaría en la misma cantidad la esperanza de vida de la clase alta y de la clase baja (lo que llamaremos política neutral), y otra política que tendría como objetivo únicamente aumentar la esperanza de vida de la clase social baja, reduciéndose así las desigualdades en salud (que llamaremos política igualitaria). Además, se administraron dos variantes de la pregunta, siendo la principal diferencia que una variante tenía un apoyo visual mientras que la otra no lo tenía. Se diseña un modelo que explica la preferencia por la política igualitaria, de acuerdo con diferentes variables demográficas, socioeconómicas e ideológicas.

Resultados: Las estimaciones probit muestran que el 70% de los entrevistados manifestaron una preferencia por la política igualitaria. Individuos de derechas o aquellos viviendo en regiones con mayor renta per capita tienen una menor probabilidad de preferir la política igualitaria. Sin embargo, sorprendentemente, ni la educación ni la renta de la familia está asociada con una mayor propensión individual a elegir la política igualitaria. Además, los jóvenes y los más viejos tienen una menor igualdad que aquellos otros de edad mediana. Finalmente, la forma en que se administra la pregunta también importa: los entrevistados que tienen un apoyo visual tienen un 20% menor de probabilidad de preferir la política igualitaria.

Conclusión: La edad, la ideología y la renta de la Comunidad Autónoma de residencia explican una actitud igualitaria frente a las desigualdades socioeconómicas en salud. De forma inesperada, la educación o la renta del hogar no parecen tener un impacto significativo. Además, la forma en que se administra la pregunta al entrevistado también cuenta: hay una probabilidad significativamente mayor de que los individuos elijan una política igualitaria cuando la gente se enfrenta a los números que a los gráficos.

EXPLORING FACTORS THAT CAN EXPLAIN PREFERENCE FOR EGALITARIAN POLICIES IN HEALTH: AN EMPIRICAL STUDY

Abstract

The main objective of this paper is to empirically explore the factors that are associated to the egalitarian attitudes regarding socioeconomic inequalities in health. A representative sample of the Spanish population was surveyed (n=1,209). After being informed that those from the higher social class have longer life expectancy at birth than those from the lower social class, respondents were required to choose between two health policies: one would increase life expectancy of the high and low social groups by the same amount (the “distribution neutral” policy), and the other option would target the lowest social class group, thereby reducing current health inequalities (the “targeting” or “egalitarian” policy). Two variants of the question were administered, the main difference being that one had a visual aid whilst the other did not. A model is developed that explains the preference for targeting according to different demographic, socioeconomic and ideological variables. Probit estimations show that 70% of respondents report a preference for the egalitarian policy. As expected, right wingers or/and individuals living in a high per capita income region are less likely to choose the egalitarian policy. However, surprisingly, neither individual’s education nor household income is associated with individual’s preference to target. Age is also a determinant of individual choice: younger and older individuals are less likely to target the egalitarian policy than those in the middle age. Finally, results also suggest that the way in which the question is asked matters: respondents that have a visual aid are less likely to target by 20%.

I. INTRODUCTION

Maximising average population health is a major objective of public health care systems. However, modern health care systems have another objective: equality in the distribution of health. Various studies have explored how members of the public see the balance between these two objectives (see for example Johannesson & Gerdtham, 1996; Anderson & Lyttkens, 1998; Cuadras-Morato et al. 2001; Dolan & Robinson, 2001; Dolan et al, 2002; Abásolo & Tsuchiya 2004). A common feature of most studies is that neither maximisation or health equalisation alone is supported, rather there is a public concern for inequalities in health to the extent that people is willing to sacrifice some gains in health in order to reduce health inequalities. However, what has been rarely investigated so far are the socioeconomic, demographic and other related characteristics which can help to explain that some people are more averse to health inequalities than others (i.e. some people have a stronger preference for egalitarian policies than others)¹. The motivation of this paper is to explore what factors can *explain* preference for egalitarian policies in health. Based on a survey data from Spain, the relationship between the respondents' backgrounds characteristics and their preferences to reduce inequalities in health is analysed. In addition, whether different modes of asking people about this issue lead to different answers is investigated.

This is done by analysing interview data obtained in Spain (see Abasolo et al. 2001). First, respondents were informed of a current socioeconomic health inequality, namely, that men from the highest social class have a longer life expectancy at birth (75 years) than men from the lowest social class (70 years)². Then, respondents were asked to choose between two health programmes with the same cost: programme A, that extends the life expectancy of both these population groups by two years each, and programme B, that targets the disadvantaged group and increases their life expectancy by four years. Since, by targeting all the benefits to the disadvantaged group, programme B will reduce the inequality by the same amount, we will refer to this as the “targeting policy” or the “egalitarian policy”. The aim of the questionnaire was to tap into the preferences of the respondent as a citizen, comparing public policies addressing issues of inequality and distribution, as opposed to a consumer, purchasing insurance policies for their own benefit. The survey question continued to identify the strength of inequality aversion by subjecting those respondents who choose the egalitarian policy to explicit trade offs between efficiency and equality; however, the present paper concentrates on the first exercise alone. At the end of the survey, socioeconomic, demographic and ideological information were gathered to serve as the explanatory variables. Two different ways of presenting the options were used (a Pictorial variant and a Verbal variant), with the aim of testing whether or not these two different devices to present the same information lead to the same results.

II. THE MODEL AND ESTIMATION METHODS

We specify a model that explains the probability of a given respondent to target the disadvantaged group, as defined above. An underlying (or latent) variable (T^*) represents an individual's propensity to chose this policy. We anticipate that demographic, socioeconomic and ideological characteristics are associated with people's attitudes towards egalitarianism. With respect to the

¹ For example, Lindholm et al (1997) shows that support for different interpretations of equity are strongly associated with sex and political affiliation.

² The data are based on those for England (Acheson, 1998). Available data on life expectancy at age 25 by education groups in Madrid and Barcelona show similar patterns (Borrel et al. 1999).

first, age (A) and gender (G) are considered. Secondly, since we are dealing with attitudes regarding socioeconomic inequalities, we may expect there to be some pattern in the responses by respondents' socioeconomic status; proxies used to explore this possibility are household income (Y), education (E), per capita income of the region of residence (R), and being in the labour market (L). In general, it is expected that all these variables are negatively related with the propensity to support the egalitarian health policy. Thirdly, political affiliation, or ideology (I), is also assumed to have a role in the model. After all, people's attitude towards egalitarianism is an ideological issue, which might be favoured more by a left wing person. Lastly, we include in the model another variable (Q) representing the questionnaire variant. The question could be presented to the respondents with the aid of pictures (P) or verbally with no visual aids (V). The aim of this regressor in the model is to determine whether there are significant differences in the estimated probabilities depending on the way the question is asked.

Thus, the model can be written as:

$$T_i^* = T(G_i, A_i, E_i, Y_i, L_i, I_i, R_i, Q_i) + \varepsilon_i \quad [\text{eq. 1}]$$

In model [eq.1], the i subscripts represent individual respondents, and ε_i captures influences unobserved by the analyst, which we assume to have a standard normal distribution with zero mean and constant variance. In practice, T^* is unobserved. Rather, we observe T_i , which is a dummy variable representing whether or not the individual actually chooses the targeting policy; therefore it is the realization of a binomial process defined by:

$$T_i = 1 \text{ if } [T_i^* > 0]$$

So, if the individual's propensity to target is positive ($T_i^* > 0$) s/he will choose the egalitarian policy ($T_i = 1$), and if otherwise ($T_i^* \leq 0$) s/he will not ($T_i = 0$).

In order to select the functional form of the empirical model, socioeconomic and statistical criteria are used. This has allowed us to consider the definitions of the set of dummy variables and whether the only continuous variable (age) enters the model in natural units, as logarithms or as higher-order powers. Interactions between regressors are also included and tested in the model. In particular, in order to test whether the mode of presenting the question (P or V) has a different effect on the propensity to target across the different background characteristics specified in the model, we will test the joint hypothesis that the interactions of Q with the rest of regressors are not significantly different from zero.

The estimation process will be undertaken through non-linear probit regressions. Likelihood ratio (LR) tests and Reset specification tests will be carried out to appraise the appropriateness of the different functional forms³.

Estimations of model [eq.1] will allow us to empirically assess the relevance of the different hypothesised explanatory variables and appraise whether the way in which the question is formulated does not account or it does, and if it does in which way.

³ Throughout, a 5% significance level is used.

Model [eq.1] may be subject to selection bias due to incomplete survey data. In surveys of this sort respondents do not always provide answers to *all* the questions of the survey. It is the so-called “item non-response”. If the pattern of non-response is not at random, conventional estimators may be biased and inconsistent. Tests for selection bias and correction, if necessary, are undertaken estimating a probit with sample selection (Greene 1997). The probit with sample selection works in a manner very similar to the Heckman model except that the response variable is binary. This method requires additional exogenous variables (or identifying variables), which should explain the probability of participating but have no direct impact on the probability to target.

So, let us assume an underlying (unobserved) variable Y_i^* that determines selection into participating groups, i.e. $Y_i = 1$ when $Y_i^* > \text{threshold}$, and $Y_i = 0$ when $Y_i^* \leq \text{threshold}$. Y_i^* would represent the inclination for the individual to participate answering all the relevant questions being Y_i^* a linear function of some of the exogenous variables in model [eq.1] as well as some identifying variables as follows:

$$Y_i^* = Y(G_i, A_i, L_i, R_i, Q_i, M_i, H_i) + u_i \quad [\text{eq. 2}]$$

The identifying variables include M_i , which represents the marital status of the individual (married, single, divorced or widow), and H_i , which represents whether the individual lives in a rural area. The main criteria used here for proposing both set of identifying variables are that the variables have an impact on the probability to participate but are unrelated to the individual’s preference for egalitarian policies. u_i is a random error term normally distributed with zero mean and constant variance.

Selection bias occurs when there is correlation between Y and ε (and therefore between ε and u); that is, when unobservable factors that influence the potentially selection are also influencing the probability to target. If so, selection bias will be corrected. To check whether selection bias is absent we will test, firstly, whether ρ (the correlation of residuals) is significantly different from zero: if the covariance between ε and u is significantly different from zero, then we cannot reject that there is no selection bias. In addition, a comparison of the estimates of both the initial probit and the probit with selection is undertaken: a large change in the coefficients, a change of the sign of the coefficients or a change in the statistical significance of the coefficients between the initial probit and the probit with selection indicate the existence of selection bias.

III. DATA AND VARIABLES DEFINITION

The data were collected during 1999 in Spain. A survey of 1,209 individuals over 18 years of age was undertaken. Face to face interviews were assigned across the 17 “Comunidades Autónomas” (Regions for short), reflecting the local resident population proportionally. Within each of the Regions, interviews were randomly allocated so that the achieved sample will be representative of the general Spanish population in terms of socio-demographic characteristics. Each of the two variants of the questionnaire was administered to approximately the same number of respondents: 602 for the P-variant and 607 for the V-variant. In terms of background characteristics, the distribution of these variables across the two variants is not statistically significantly different from each other. For descriptive statistics, see Table 1 of the Appendix.

The binary dependent variable, *target*, takes the value 1 if individual i targets the worse off group, 0 if otherwise. *Age* is a continuous variable, which takes values between 18 and 94, and also enters

the equation in quadratics. The binary variable *female* indicates whether the individual is female or male. Education is recorded by level of schooling and has been categorised in three dummy variables representing low education *lowedu* (those with primary school education or less), middle education *midedu* (those with secondary school education, the baseline category), and high education *highedu* (those with higher and university education). Household income is also categorised in three dummies *high income* (more than 1,653 euros per month), *middle income* (between 600 and 1,653 euros per month, the baseline category) and *low income* (less than 600 euros per month)⁴. Per capita income in the region of residence is captured by three dummies: high income regions *highreg* (those resident in Madrid, Navarra or País Vasco), low income regions *lowreg* (those who live in Andalucia or Extremadura) and middle income regions *midreg* (residents in the rest of Spain, the omitted category). The binary variable non-labour market (*nonlabmkt*) indicates whether the individual is not currently in the labour market (i.e. is retired, unemployed, homemaker or student) or whether s/he is. Political affiliation is recorded by three categorical indicators, *right* (those who report as being centre-right, right or extreme right wing), *left* (those who report as being centre-left, left or extreme left wing), and *centre* (those who are in the political centre, the baseline category). *Type* is the dummy variable representing the type of question administered, taking value 1 if the V variant was provided, and 0 if the P variant was administered. Regarding the identifying variables, *single*, *divorced* and *widow* are three dummy variables representing whether the individual belongs to one of such marital status (as opposed to the omitted category *married*); and the binary variable *rural*, representing whether the individual is resident in a rural area (of less than 10,000 residents).

IV. RESULTS

Descriptive statistics and estimation results are reported in table 1. Overall, 70% of respondents chose to target the egalitarian policy. Model [eq.1] passes the Reset specification test, indicating that there is no evidence of functional form problems (table 2). In addition, when dealing with the issue of the relevance of the question variant in the model, likelihood ratio tests show that specification [eq.1] fits the data significantly better than other alternative specifications. The alternative equations considered include, on the one hand, a model that omits the *type* variable and, on the other hand, a model which includes the *type* variable together with interactions with the rest of independent variables. Therefore model [eq.1] shows that the way in which the question is presented counts, although there are no differential effects across the individual characteristics considered in the model.

Item non-response leads to 514 missing cases, which corresponds to 42.5% of the entire data, leaving 695 individuals as valid cases. Estimates for the probit with sample selection [eq.2] can be seen in table 4. The correlation coefficient (*rho*) is not statistically different from zero. It is also true that the confidence interval for the parameter estimate is wide. However, sign, magnitude and t-ratios of coefficients of the probit with selection are very similar to those of the initial probit estimation [eq.1]. Overall, the results suggest that we cannot reject the null hypothesis that there is no selection bias.

Probit average and marginal effects evaluated at sample means (table 3) show that: firstly, other things equal, age is a statistically significant explanatory variable of the probability to target the disadvantaged group. However, the marginal effect of age on targeting is not constant and changes with the age of the respondent. For young adults between 18 and 44 the probability of targeting

⁴ Although it would have been desirable to derive equivalent income, this has not been possible given the available information in the survey and also given that the variable household income was not available as a continuous variable.

increases with age at a diminishing rate. A maximum is reached, on average, at the age of 44, after which the probability of targeting starts to decrease with age at an increasing rate.

Secondly, those who are politically right wing have a significantly lower probability of targeting the egalitarian policy, as compared to those who are in the political centre (and consequently, as compared also to those who are left wing). In particular, the probability of a right wing individual targeting is on average 12% less than the reference individual, other things being equal. Finally, living in one of the richest regions such as Madrid, Navarra or País Vasco is associated with a lower probability of targeting the egalitarian policy, other things being equal, by 17.4%, compared to those living in average per capita income regions. There is no evidence that other demographic and socioeconomic factors included in the model have a significant influence in the attitudes towards health inequalities.

Regarding the effect of the question variants, our model suggests that mode of presentation makes a difference. When individuals are administered the V-variant (as opposed to the P-variant), there is a 7% higher probability of targeting the egalitarian policy (at sample means). A comparison of this average effect with those of other variables shows that, at sample means, the average effect of the question variants is higher than that the marginal effect of age but lower than the average effect of being right wing or resident in a rich region. However, there is no evidence that the question variant has differential effects on the probability of targeting across the different demographic, socioeconomic and ideological characteristics, when these are taken jointly.

V. DISCUSSION

The fact that a vast majority of individuals in society are averse to health inequalities and show a preference for an egalitarian health policy, raises the issue of *how* this happens. This is explored with reference to the predicted probability to target of a “standard Spaniard”: an individual who is of average age (45 years old) and belongs to the baseline categories of our dummy variables, that is, male, with middle income and education levels, who is in the labour market, politically centred, resident in a middle per capita income region and has been administered the P-variant of the question. This standard Spaniard has a 68% predicted probability of choosing the targeting policy. That is, two thirds of those with our standard features would support that resources ought to be devoted to the lowest (and unhealthiest) socioeconomic group.

The results suggest that age has a role when explaining this (see figure 1): at age 18, the predicted probability that the standard Spaniard targets is 56%. Starting from here, it increases with age, reaching the maximum probability of 69% at age 44, after which this probability diminishes with additional years. This predicted profile is symmetric with respect to age so that, for example, probabilities at age 24 (= 44 - 20) and at age 64 (= 44 + 20) are both 61%. The age at which the predicted probability of targeting is 50% is age 74. Unfortunately, it is impossible to tell from our data whether this is an age effect (different propensity to target depending on one’s stage in life) or a cohort effect (different propensity to target depending on one’s birth cohort), or a mixture of both. So, with no more information, it is difficult to highlight the reasons behind this age pattern in attitudes towards health inequalities.

Those who are politically right wing have a significantly lower probability of choosing the egalitarian policy, as compared to those of the centre (and consequently as compared to those left wing). Therefore, if the standard Spaniard became right wing (other things remaining equal) this would reduce the probability of targeting by 13%. This is consistent with prior expectations, that

egalitarianism is associated with left wing ideology rather than right wing. However, the estimated average effect might be overestimated if, as it seems, right wing people are more reluctant to explicitly indicate their political affiliation in the background questions⁵. Finally, if a standard Spaniard was resident in one of the richest Regions like Madrid, Navarra or País Vasco (instead of living in a middle income Region) there would be, on average, a 18% lower probability of targeting the egalitarian policy.

Of the other demographic and socioeconomic factors with possible association with respondents' propensity to target, none were statistically significantly associated. Particularly, socioeconomic background (approximated by education and household income variables) does not seem to affect inequality aversion. Rather, health inequalities seem to be a matter of concern for the general population, irrespective of their income or education level. Taking into account that we are dealing with socioeconomic inequalities in health, these results are not consistent with selfish motivations. To the contrary, they seem to have picked up the respondents playing the role as citizens facing a policy rather than the role of consumers purchasing insurance policies, which was the intention of the questionnaire.

The way in which the question is presented is associated with the probability of a given respondent choosing the egalitarian policy. People have a higher propensity to target when they deal with the verbal variant than when they face the pictorial variant. A possible explanation for this is that the pictorial variant may have emphasised the gain more than the existing inequality. Imagine a respondent who regards him/herself as an "egalitarian" person, who cares about equality. The definition of egalitarianism usually used in economic theory is consequence-based, as opposed to gain-based, but this might not agree with lay perception. In other words, contrary to the economists' conceptualisation, programme A, which gives equal amounts to both groups, might be perceived by the lay respondent as the egalitarian alternative, rather than programme B, which gives nothing to one group and all the benefits to the other. The illustrated P-variant seems to be at a higher risk of this happening, because it visualises the differences across the two programmes in terms of what the two population groups will gain. The current baseline inequality is the same across the two programmes, so they could be cancelled out from the cognitive process. Should this happen, then those respondents who, had they been given the V-variant may have chosen to target, may well choose not to target in the P-variant *precisely because they regard themselves as egalitarian*. Although the current data do not lead us to discriminate in favour of one particular mode of administration, we believe that this issue deserves more research in order to explore the reasons that lead people to choose differently under these two different formats.

It should also be noted that, despite this variant effect, the overall proportion of respondents who chose to target is roughly the same across the two questionnaire variants and the difference is not statistically significant. In certain research contexts, where the aim of the study is not to explain the variation in targeting behaviour by respondent background characteristics, this may be seen as sufficient justification to pool the data from the two variants. However, what the present analysis has demonstrated is that although the null hypothesis that the two questionnaire variants produce the *same* aggregate results cannot be rejected on the basis of statistical significance, the composition of respondents who make up these aggregate results are different. In this respect, the two variants are *different*, and the agreement at the aggregate level is a chance event with no meaning.

⁵ Only 21% of individuals report that they are centre-right or right, whilst 47% of respondents report a centre-left or left ideology. In 1999, the Spanish Popular Party was in government (a centre-right party) and the survey was undertaken just 3 months before general elections which the Popular Party won again with absolute majority.

VI. CONCLUSION

A survey that elicited peoples' aversion to inequality in health was carried out in Spain. The current study explored the factors associated with the respondents' propensity to choose a policy that reduced the inequality, as opposed to a policy that did not. These included the respondents socioeconomic and demographic characteristics, their political affiliations, and the mode of presentation used in the questionnaire. The results suggest that age, ideology and the affluence of the region where they live explain people's aversion to inequality in health: older individuals, right wingers and/or those living in affluent regions are significantly less likely to support egalitarian policies. Unexpectedly, other socio-economic background characteristics like individual's education or household's income do not seem to have a significant impact. The evidence also shows that the way in which the question is administered also counts: the regression results indicate that on average, there is a significantly higher probability to target when people deal with numbers rather than with pictures.

VII. REFERENCES

Abasolo I., Shaw R., Tsuchiya A. (2001) ¿Hasta qué punto están más preocupados los españoles por la equidad que por la eficiencia en salud? Comunicación presentada en las XXI Jornadas de la Asociación de la Economía de la Salud, 6-8 June, Ovideo.

Abásolo I. y Tsuchiya, A. (2004) Exploring Social Welfare Functions and violation of monotonicity: An example from inequalities in health. *Journal of Health Economics*, 23 (2): 313-329.

Acheson D. (1998) *Independent Inquiry into Inequalities in health*. London, HMSO.

Andersson F. y Lyttkens CH. (1999). Preferences for equity in health behind a veil of ignorance. *Health Economics*;8(5):369-78.

Borrell C., Regidor E., Arias L. (1999) Inequalities in mortality according to educational level in two large southern European cities. *International Epidemiological Association*, 28: 58-63.

Cuadras-Morato X, Pinto-Prades JL, Abellán-Perpiñan JM (2001). Equity considerations in health care: The relevance of claims. *Health Economics*;10(3):187-205.

Dolan P. y Robinson A (2001). The measurement of preferences over the distribution of benefits: The importance of the reference point. *European Economic Review*;45(9):1697-709.

Dolan P, Tsuchiya A, Smith P, Shaw R, Williams A. (2002) *Determining the parameters in a social welfare function using stated preference data: an application to health*. Sheffield Health Economics Study Group, University of Sheffield. Discussion Paper Series 02/2.

Greene WH (1997) *Econometric Analysis*. Prentice Hall, third edition. New Jersey.

Johannesson M, y Gerdtham UG (1996). A note on the estimation of the equity-efficiency trade-off for QALYs. *Journal of Health Economics*;15:359-368.

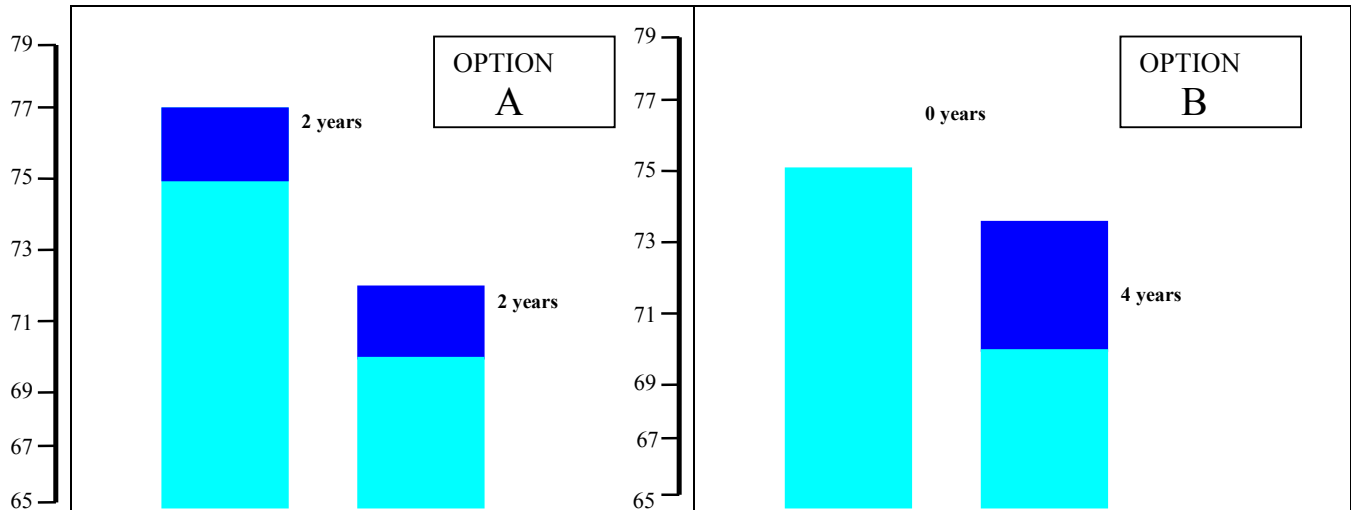
Lindholm, L., Emelli, M. and Rosén, M. (1997) Health maximization rejected -the view of Swedish health care politicians. *European Journal of Public Health*; 7: 405-10.

Wagstaff A (1994). QALYs and the equity-efficiency tradeoff. In: Layard A, Glaister S, eds. *Cost-Benefit Analysis*. second edition: Cambridge University Press (reprinted from *Journal of Health Economics*, 10:21-41, 1991 with corrections)

Williams A (1997). Intergenerational equity: an exploration of the 'fair innings' argument. *Health Economics* 6: 117-132

APPENDIX

OPTIONS PRESENTED IN THE PICTORIAL VERSION (P)



OPTIONS PRESENTED IN THE VERBAL VERSION (V)

<p>Increase high class life expectancy in 2 years and low class life expectancy in 2 years. The final outcome of life expectancy would be: high class: 77 years low class: 72 years</p>	<p>OPTION A</p>
<p>Increase only low class life expectancy. The final outcome of life expectancy would be: high class: 75 years low class: 74 years</p>	<p>OPTION B</p>

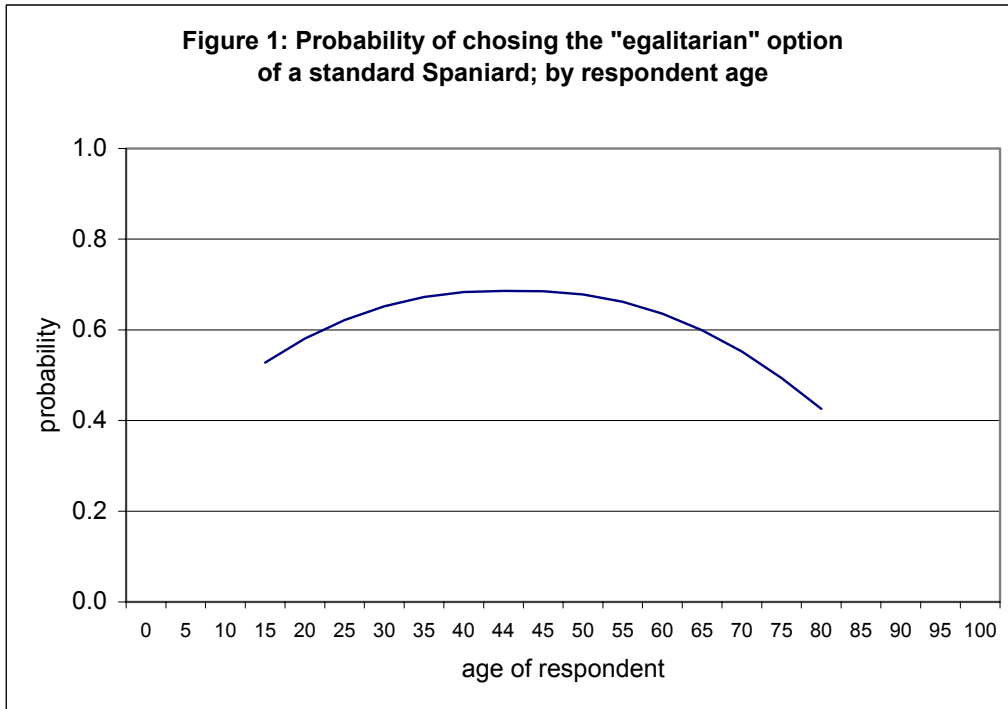


TABLE 1. MEAN VALUES AND STANDARD DEVIATIONS OF VARIABLES

Variable	Obs	Weight	Mean	Std. Dev.	Min	Max
target	695	694.516207	.6904674	.4626339	0	1
female	695	694.516207	.485911	.5001614	0	1
age	695	694.516207	45.26725	17.5467	18	91
agesq	695	694.516207	2356.567	1709.339	324	8281
highedu	695	694.516207	.2362348	.425074	0	1
midedu	695	694.516207	.4139528	.492895	0	1
lowedu	695	694.516207	.3498124	.477254	0	1
highinc	695	694.516207	.2259842	.4185301	0	1
midinc	695	694.516207	.4759503	.499781	0	1
lowinc	695	694.516207	.2980655	.4577378	0	1
nonlabmkt	695	694.516207	.5352979	.4991117	0	1
left	695	694.516207	.4718532	.4995667	0	1
centre	695	694.516207	.3091312	.4624682	0	1
right	695	694.516207	.2190156	.4138771	0	1
reghigh	695	694.516207	.1453269	.352684	0	1
regmid	695	694.516207	.6314791	.482751	0	1
reglow	695	694.516207	.223194	.4166872	0	1
type	695	694.516207	.50066	.5003597	0	1

TABLE 2. RESULTS PROBIT ESTIMATION

```
. probit target female age agesq highedu lowedu highinc lowinc nonlabmkt left
right reglow reghigh type [aweight=weight]
```

```
(sum of wgt is 6.9452e+02)
Iteration 0: log likelihood = -430.0147
Iteration 1: log likelihood = -411.00042
Iteration 2: log likelihood = -410.96762
Iteration 3: log likelihood = -410.96762
```

```
Probit estimates                               Number of obs   =           695
                                                LR chi2(13)    =           38.09
                                                Prob > chi2    =           0.0003
Log likelihood = -410.96762                    Pseudo R2      =           0.0443
                                                RESET test 0.04;  sq(1)=3.84
```

target	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
female	.0356847	.1081395	0.33	0.741	-.1762649 .2476343
age	.0439998	.0173643	2.53	0.011	.0099663 .0780333
agesq	-.0005051	.0001825	-2.77	0.006	-.0008628 -.0001475
highedu	.2193948	.1342858	1.63	0.102	-.0438006 .4825902
lowedu	.2170204	.1467327	1.48	0.139	-.0705704 .5046111
highinc	-.1167251	.1338214	-0.87	0.383	-.3790103 .14556
lowinc	.0217484	.136658	0.16	0.874	-.2460964 .2895933
nonlabmkt	.1568169	.1272843	1.23	0.218	-.0926558 .4062896
left	-.0148879	.1222525	-0.12	0.903	-.2544983 .2247225
right	-.3359247	.1401156	-2.40	0.017	-.6105463 -.0613031
reglow	.1271331	.1309688	0.97	0.332	-.1295612 .3838273
reghigh	-.4672769	.1422487	-3.28	0.001	-.7460792 -.1884745
type	.2037721	.1021195	2.00	0.046	.0036215 .4039227
_cons	-.4742725	.4152314	-1.14	0.253	-1.288111 .3395662

. RESET TEST

```
. quietly probit target $xvars targetf2
```

```
. test targetf2=0
```

```
( 1) targetf2 = 0.0
```

```
chi2( 1) = 0.30
Prob > chi2 = 0.5835
```

TABLE 3. PROBIT MARGINAL AND AVERAGE EFFECTS

Probit estimates

Number of obs = 695

LR chi2(13) = 38.09

Prob > chi2 = 0.0003

Pseudo R2 = 0.0443

Log likelihood = -410.96762

target	dF/dx	Std. Err.	z	P> z	x-bar	[95% C.I.]
female*	.0124443	.0376975	0.33	0.741	.485911	-.061442	.08633	
age	.0153486	.0060523	2.53	0.011	45.2672	.003486	.027211	
agesq	-.0001762	.0000636	-2.77	0.006	2356.57	-.000301	-.000052	
highedu*	.0740483	.0436885	1.63	0.102	.236235	-.01158	.159676	
lowedu*	.0742953	.049191	1.48	0.139	.349812	-.022117	.170708	
highinc*	-.041359	.0481081	-0.87	0.383	.225984	-.135649	.052931	
lowinc*	.0075692	.0474503	0.16	0.874	.298065	-.085432	.10057	
nonlab~t*	.054818	.0445468	1.23	0.218	.535298	-.032492	.142128	
left*	-.0051945	.0426634	-0.12	0.903	.471853	-.088813	.078424	
right*	-.1220372	.0524927	-2.40	0.017	.219016	-.224921	-.019153	
reglow*	.0434998	.0438869	0.97	0.332	.223194	-.042517	.129516	
reghigh*	-.1738084	.0550714	-3.28	0.001	.145327	-.281746	-.06587	
type*	.0709977	.0354917	2.00	0.046	.50066	.001435	.14056	
obs. P	.6904674							
pred. P	.6978101	(at x-bar)						

(*) dF/dx is for discrete change of dummy variable from 0 to 1
z and P>|z| are the test of the underlying coefficient being 0

TABLE 4. PROBIT WITH SAMPLE SELECTION

Number of obs = 1209
 Censored obs = 514
 Uncensored obs = 695
 Wald chi2(13) = 38.00
 Prob > chi2 = 0.0003
 Log likelihood = -1205.962

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

target						
female	.0057774	.1480517	0.04	0.969	-.2843985	.2959534
age	.0489134	.0217279	2.25	0.024	.0063275	.0914993
agesq	-.0005496	.000214	-2.57	0.010	-.0009691	-.0001302
highedu	.2097602	.1384156	1.52	0.130	-.0615294	.4810498
lowedu	.2111436	.1451737	1.45	0.146	-.0733917	.4956788
highinc	-.1141935	.1327207	-0.86	0.390	-.3743213	.1459343
lowinc	.0226788	.1338767	0.17	0.865	-.2397148	.2850724
nonlabmkt	.1352117	.1500073	0.90	0.367	-.1587972	.4292207
left	-.0149317	.1195964	-0.12	0.901	-.2493362	.2194729
right	-.3305266	.1424466	-2.32	0.020	-.6097169	-.0513364
reglow	.1492482	.1469055	1.02	0.310	-.1386813	.4371778
reghigh	-.534301	.2321363	-2.30	0.021	-.9892797	-.0793222
type	.2003553	.1029275	1.95	0.052	-.0013789	.4020894
_cons	-.7403945	.9407205	-0.79	0.431	-2.584173	1.103384

Y						
female	-.2007844	.0792574	-2.53	0.011	-.356126	-.0454427
age	.0364361	.0136711	2.67	0.008	.0096411	.063231
agesq	-.0003817	.0001372	-2.78	0.005	-.0006505	-.0001129
nonlabmkt	-.0980364	.088269	-1.11	0.267	-.2710405	.0749677
reghigh	-.4232526	.0963924	-4.39	0.000	-.6121783	-.2343269
reglow	.1806764	.0976276	1.85	0.064	-.0106701	.372023
single	-.1022373	.1099509	-0.93	0.352	-.317737	.1132624
divorced	-.1270649	.2394361	-0.53	0.596	-.596351	.3422211
widow	.2583478	.152567	1.69	0.090	-.0406781	.5573737
rural	.1091552	.0928946	1.18	0.240	-.0729148	.2912253
_cons	-.3541731	.3366173	-1.05	0.293	-1.013931	.3055846

/athrho	.2713856	.9422371	0.29	0.773	-1.575365	2.118136

rho (ρ)	.2649136	.8761116			-.917875	.9714895

LR test of indep. eqns. (rho = 0):	chi2(1) =	0.08	Prob > chi2 =	0.7742		
