

Paper presented to HESG Conference, Oxford, January 1-3, 2001

*Not for citation without authors permission –
What follows is an example of work that is very much in progress.*

Some methodological issues in the measurement of people's preferences regarding age weighting in health

Paul Dolan, Universities of Sheffield and Oslo (P.Dolan@shef.ac.uk)

Rebecca Shaw, University of York (rls5@york.ac.uk)

Aki Tsuchiya, University of Sheffield (A.Tsuchiya@sheffield.ac.uk)

1. INTRODUCTION

One of the attributes that may be taken into account when setting priorities in health care is the age of the potential recipients. Let us refer to differential treatment by age as “ageism”.¹ In this paper, we consider some of the methodological issues that are involved in trying to provide quantitative information about people’s preferences over ageism in health care. We also present the results from our own empirical study that was designed to shed some light on these issues.

1.1 Concepts of ageism

According to Tsuchiya (1999), there are three main types of ageism. The first will be referred to as “QALY maximisation ageism” (QMA).² *Ceteris paribus*, QMA will give priority to a younger person over an older one since the former will usually experience any health gains for longer. QMA is compatible with the assumption that each QALY is of equal value, irrespective of who receives those QALYs. Therefore, the central concept in QMA is not age *per se*; rather it is life expectancy, which will depend heavily (but not exclusively) on current age.

The second type of ageism is “productivity ageism” (PA). This gives priority to young adults because they are more productive at home and in society.³ A QALY at different ages will then be valued differently according to the expected level of productivity at each age. The age weights used in the calculation of the Burden of Disease (Murray, 1996) is an example of this type of PA. The third type of ageism is “fair innings ageism” (FIA) which looks at the number of QALYs people can expect

¹ In lay language, this term may suggest unacceptable discrimination on the basis of age but, in this paper, the term is neutral with regard to the legal, ethical and political appropriateness of age weighting.

² This type of ageism is sometimes referred to as “utilitarian ageism” (Nord *et al.*, 1996), where “utilitarianism” is understood in a very narrow and specific way.

³ “Productivity” here refers to all kinds of activities, and should not be understood in a narrow market-oriented way.

to have over their lifetime.⁴ FIA will give priority to a younger person over an older one because, *ceteris paribus*, the former can expect a smaller number of lifetime QALYs than the latter (Williams, 1997).

1.2 Studies of ageism

There have been several attempts to study people's preferences about age weighting in health (for a review of some of these, see Tsuchiya, 1999). Table 1 provides a classification of the studies according to the types of questions that have been asked. Some studies appear in more than one cell because more than one kind of question has been employed within the study. The usual approach involves asking respondents which of two patients or groups should be given priority over the other when it is not possible to treat both at once. Typically, respondents are told that the two patients or groups differ only by age.

Whether and how the studies have tried to quantify the extent to which one party is preferred over the other is shown in the columns of the table. Some studies have elicited only ordinal preferences, whilst those that have attempted to quantify preferences have done so in two ways. One has been to use a person trade-off (PTO) design, and to change the number of people in the two groups in order to make the respondent indifferent between the groups. The other has been to change the size of the health benefits that each patient or group can expect until a point of indifference is reached. The rows represent the different ways in which health benefits have been expressed. Benefits have either been in terms of life-saving or health improvement, and these have been described as lasting for the rest of the recipients lives or for a fixed period. To our knowledge, no study has presented benefits in terms of lifesaving and health improvement.

In general, the empirical evidence suggests that the age weighting function declines after middle age and peaks either in middle age or in childhood. These results do not support QMA and, depending on where the peak is, provide some support for PA and FIA.⁵ It is difficult, then, from the studies to date to distinguish fully between the different types of ageism. Moreover, very few of them have attempted to generate an overall age weighting function.

1.3 Methodological issues

Table 1 suggests that there are a number of considerations when deciding upon which question format to use. First, let us deal with the columns in Table 1. In order to elicit preferences with a view to generating age weights, those preferences clearly need to be quantified. There would seem to be no obvious reasons for choosing to quantify age weights in terms the *number of people* in one group vis-à-vis another, or in terms of the *benefits* to one group vis-à-vis another. Ultimately, the choice between them would be determined by the extent to which value is linear in the number of people in the former case and linear in health benefits in the latter case.

⁴ Nord *et al* (1996) refer to this as "egalitarian ageism" but egalitarianism does not need to be based upon a fair innings argument.

⁵ These results have been found to be largely independent of the respondent's own age and other background characteristics.

Now consider the rows in Table 1. There are certainly difficulties with expressing benefits over the remaining lifetime (i.e. the second and fourth rows). The responses to questions where benefits are expressed in this way will have to be decomposed to derive weights for individual years of life at different ages. However, the responses to such questions will be a function of: 1) the life expectancy of the recipients as perceived by the respondent; and 2) the rate at which the respondent discounts health benefits. Since neither of these can be observed directly, it would seem to make more sense to use a fixed duration of health benefit; that is, to use the question formats on the third and fifth rows of Table 1.

We can perhaps narrow our recommendations down still further since there is a complication in eliciting preferences in terms of health improvements. The standard assumption in the literature on QALYs is that the decrement in quality of life due to a given health problem is independent of the person's age. For example, if having a specific problem in mobility is to count as a 10% loss in quality of life, then it will count as 10% loss to a 15 year-old and to a 75 year-old. But how many respondents would really believe this? We suspect that they would be much more likely to think that a given problem in mobility would represent a larger loss to the 15 year-old than to the 75 year-old. In other words, the loss in quality of life from a given health state will vary – in the respondent's mind – according to the age of the person experiencing that loss, and so preferences will be confounded by this.

Based on these considerations, the questions in this study elicit preferences using the PTO method where benefits are expressed in terms of a five-year extension in life. Of course, this format is not free from some of the problems highlighted above e.g. respondents may still consider the *value* of a five-year benefit to be a function of the age of the recipient. However, we suggest that they are likely to be less. In any event, by collecting qualitative data about the reasons for particular responses, we are able to provide some insights into the extent of the problems.

2. METHODS

2.1 The ranking exercise

A questionnaire was designed during a period of qualitative research involving in-depth interviews and focus groups and was refined during extensive piloting. There were two rounds of interviews. In the first round, respondents were presented with five groups of different ages, selected to represent five different stages in the life cycle. The five ages were 5 year-olds, 20 year-olds, 35 year-olds, 55 year-olds and 70 year olds. Respondents were told that each group would die in a few days. They were then asked to rank the groups in the order in which they would choose to give an extra five years of life. They were probed to explain their reasoning using a set of pre-determined categories that were generated during the pilot interviews.

Respondents in the second round were presented with exactly the same question – but only after they had ranked the same five ages on the basis that they could prevent the premature deaths of people who would otherwise die in a few days but who would live a normal life expectancy if treated. Respondents were again probed for their reasoning. The objective of this question was to highlight the limited nature of the

(five-year) benefit in the next question by opposing it to a benefit that lasted for the remainder of the recipient's life.

The chi-square test was used to look at the relationship between the rankings and the reasons for those rankings. Each background variable was recoded into two groups and univariate analysis (again using the chi-square) was performed to look at whether the rankings or the reasons for them were affected by any respondent characteristics. In order to test whether the ranking of each age in the 'five year benefit' question was affected by being asked a 'full life benefit' question first, a chi-square test was used to test for differences in the distribution of ranks for each age across the two rounds of interviews. The same test was used to test for differences in the reasons for the rankings across the two rounds.

2.2 The PTO questions

After the ranking exercise(s), all respondents were asked three PTO questions that asked them to consider pairs of programmes that differed with respect to the age of the people involved. The age groups used were those that the respondent had ranked first, third and fifth in the previous question. They were asked to indicate how many people there would have to be in the second programme relative to the first for the two programmes to be of equal social value. The precise format of the PTO questions is shown in the Appendix. Respondents were again probed to explain their reasoning. In the first round of interviews, the starting point was 100 people in each age group, and in the second round it was 1000 people.⁶

The three PTO questions were chosen to test the ordinal and cardinal properties of the responses. Since the distance, in terms of strength of preference, between the 5th and 1st ranked ages should be greater than each of the distances between the 5th and 3rd and 3rd and 1st ranked ages, there are two consistency conditions that each respondent's PTO responses should satisfy. First, the number of 1st ranked ages that is equivalent to x 5th ranked ages should be less than the number of 3rd ranked ages that is equivalent to x 5th ranked ages and, second, it should also be less than the number of 1st ranked ages that is equivalent to x 3rd ranked ages. This should hold for all values of x . In other words, the marginal rate of substitution (MRS) between the fifth and first ranked ages should be greater than the MRS between the 5th and 3rd ranked, and 3rd and 1st ranked, ages i.e. $MRS_{5v1} > MRS_{5v3}$ and $MRS_{5v1} > MRS_{3v1}$. To determine whether there were any ordering effects, half the respondents in each round of interviews received the order MRS_{5v1} , MRS_{5v3} , then MRS_{3v1} , and half received the order MRS_{3v1} , MRS_{5v1} , then MRS_{5v3} .

Researchers have typically interpreted responses to PTO questions as the *ratio* of the number of people in the two groups. If our respondents are basing their PTO responses on the ratio between the numbers of people in the two ages, then $MRS_{5v3} * MRS_{3v1} = MRS_{5v1}$.⁷ However, $MRS_{5v3} * MRS_{3v1} > MRS_{5v1}$ would suggest that respondents are also concerned about the *difference* between the number of people in the two groups, as well as about the ratio between them. If respondents do focus on the difference in the number of people in the two groups, then the difference between

⁶ The results from the pilot interviews demonstrate that changing the numeraire does not affect the ordering of preferences over the age groups.

⁷ There seems little reason to suppose that $MRS_{5v3} * MRS_{3v1} < MRS_{5v1}$.

the implied and actual MRS_{5v1} will increase as the baseline number of people in the PTO question increases from 100 to 1000.

2.3 The sample

Every 8th person appearing on the electoral register in three wards in York was contacted and invited to participate. Out of a total of 1,500 letters initially distributed, 467 people (31%) agreed to participate. To ensure that the sample was broadly representative of the wider population, 140 respondents were selected for interview based on information on a broad range of characteristics obtained from their reply slips. In total, 130 individuals were interviewed by one of the authors and two other researchers. The interviews took place at the University of York and lasted for about one hour, of which twenty minutes was spent on the questions analysed in this paper. Table 2 shows the characteristics of the sample – and how it compares quite favourably with that of Yorkshire and Humberside – and Table 3 shows the number of respondents in each of the experimental cells.

3. RESULTS

3.1 The ranking exercise

Table 4 shows the percentage of respondents in the two rounds of interviews according to where they ranked a given age group. Thus, for example, 53% of respondents in round one ranked 5 year-olds first, while 39% from round two did the same. While there is a small difference in the ranking of 70 year olds across the rounds, there is a large discrepancy between the rounds for the ranking of ages 5 and 35. This is highlighted in the significant changes to the mean ranks across the two rounds, as shown in Table 5.

Table 6 shows the reasons for the rankings by round of interview. There are striking differences across the two rounds – differences that have important implications for the concepts of ageism used. From the results of round one, it would appear that respondents are overwhelmingly in favour of FIA, with two-thirds of respondents citing ‘lived less life’ as the main reason for the responses. In round two, however, this percentage half as much larger number cite ‘five years is more valuable to some’ and a few more respondents opt for ‘family responsibilities etc.’ Responses in round two, then, would appear to be driven much more by PA.

In terms of the relationship between the rankings and reasons for them, those who have 35 year-olds ranked first are more likely to give ‘family responsibilities etc.’ as the reason for their ranking and those who have 70 year-olds ranked last are more likely to give ‘lived less life’ as the reason. These results, of course, have intuitive appeal. There is very little evidence that the rankings or the reasons for them are related to particular respondent characteristics, including age. The only differences that emerge are that females in round one have 35 year-olds ranked more highly and are more likely to give family responsibilities etc. as the reason for their ranking. These differences completely disappear in round two.

3.2 The PTO questions

Table 7 shows the results from the consistency tests. There does not appear to be any systematic pattern in these consistency rates. Round one is more consistent than round two on the first consistency condition and order two is more consistent than order one on the second consistency condition. Overall, this means that highest consistency rates are found in round one order two and lowest consistency rates are found in round two order one. The reasons for this are not at all clear and may just simply be due to the particular respondents in these cells.

Figures 1 and 2 shows each individual's actual response to the "5v1" PTO question plotted against the response that is predicted from his/her responses to the "5v3" and "3v1" PTO questions. Notice that the majority of respondents have an implied MRS that is greater than the actual one; that is, they are willing to trade off *less* first ranked for fifth ranked people than is implied from their responses to the two other PTO questions. Notice also that this trend is stronger for round two (where the numeraire is 1000 people) than it is for round one (where the numeraire is 100 people). This is precisely the pattern that would be expected if people paid some attention to the absolute difference in the number of people in the two groups.

Table 8 shows the three ages that were used in the PTO questions; that is, the ages that were ranked first, third and fifth by each respondent. Further analysis is required to determine whether the MRS of one age for another is a function of the precise ages used but early analysis suggests that this is not the case. In other words, the responses to the PTO questions do not appear to be related to the particular ages used.

4. DISCUSSION

There are three main findings from this study, each of which raises serious questions about the extent to which it is possible to derive 'pure' age weights:

1. First, presenting respondents with a 'full life benefit' question before a 'five year benefit' question has a profound effect on their responses. As evidence of this, a number of respondents stated that the shift in their ranking of the 35 year olds was directly due to exposure to the 'full life benefit' e.g. *"For five years the choice is more difficult. It's only five years, so you have to think about who would get most from an extra five years and the thirty-five year old probably has young children"*. The important methodological question here is, if you want people to answer the question *you* have in mind, first ask them the question you think *they* may have in mind.
2. Second, preferences over ages are a function of the life-stage people are seen to be at. A number of respondent quotes illustrate this. Some were indicative of PA e.g. *"The thirty-five year old has more to lose. Their death would leave a lot of people bereft. The twenty year old is probably on the first step of their career ladder, but is single and has no responsibilities"*. Others were more supportive of FIA e.g. *"Death at a young age is a tragic event. At seventy you've had your three score and ten, you're aware of your mortality and may perhaps die at any time anyway. It's less shocking if a seventy year old dies, less tragic"*.

3. Third, PTO responses are driven by the relative *and the absolute* difference in the number of people in the two groups. This means that responses to such questions cannot automatically be interpreted in the way that analysts would ideally like to interpret them.

REFERENCES

- Busschbach, J.J.V., Helsing, D.J., de Charro, F.T. (1993), The utility of health at different stages in life: A quantitative approach, *Social Science and Medicine*, 37(2):153-158
- Charny, M.C., Lewis, P.A., Farrow, S.C. (1989), Choosing who shall not be treated in the NHS, *Social Science and Medicine*, 28(12):1331-1338
- Cropper, M.L., Aydede, S.K., Portney, P.R. (1994), Preferences for life saving programs: How the public discounts time and age, *Journal of Risk and Uncertainty*, 8:243-265
- Johannesson, M., Johannsson, P-O. (1997), Is the valuation of a QALY gained independent of age? - Some empirical evidence, *Journal of Health Economics*, 16: 585-599
- Murray, C.J.L. (1996), Rethinking DALYs. In Murray, C.J.L. and Lopez, A.D. (eds.), *The Global Burden of Disease* : Harvard University Press
- Nord, E., Richardson, J., Street, A., Kuhse, H., Singer, P. (1996), The significance of age and duration of effect in social evaluation of health care, *Health Care Analysis*, 4:103-111
- Nord, E., Richardson, J., Street, A., Kuhse, H., Singer, P. (1995), Maximizing health benefits versus egalitarianism: An Australian survey of health issues, *Social Science and Medicine*, 41(10):1429-1437
- Patrick, D.L., Bush, J.W., Chen, M.M. (1973), Methods for measuring levels of well-being for a health status index. *Health Services Research*: 228-245
- Ratcliffe, J. (2000), Public preferences for the allocation of donor liver grafts for transplantation, *Health Economics*, 9: 137-148
- Rodríguez, E., Pinto, J.L. (2000), The social value of health programmes: Is age a relevant factor?, *Health Economics*, 9:611-621
- Tsuchiya, A. (1996), The value of health at different ages, *Journal of Health Care and Society*, 6(3):123-136, (in Japanese)
- Tsuchiya, A. (1999), Age-related preferences and age weighting health benefits. *Social Science and Medicine*, 48(2): 267-276
- Williams, A. (1997), Intergenerational equity: An exploration of the 'fair innings' argument. *Health Economics*, 6: 117-132

Table 1: Classification of empirical studies of ageist preferences

Quantification			
The benefit	unquantified	quantified in terms of number of people	quantified in terms of the benefit
life saving, for remaining life	Charny <i>et al</i> , 89 Nord <i>et al</i> , 95 Nord <i>et al</i> , 96 <i>This study</i>	Cropper <i>et al</i> , 94 Johannesson <i>et al</i> , 97	
life saving, for fixed period		Nord <i>et al</i> , 96 <i>This study</i>	Rodríguez and Pinto, 00 Ratcliffe, 00
health improvement, for remaining life	Nord <i>et al</i> , 95		
health improvement, for fixed period		Nord <i>et al</i> , 96	Busschbach <i>et al</i> , 93 Tsuchiya, 96

Table 2: Respondent characteristics

Category	Round 1	Round 2	York and Humberside
Sex:			
Male	48%	49%	47%
Female	52%	51%	53%
Age:			
<44 years	48%	51%	49%
>45 years	52%	49%	51%
Mean age	50	48	
Dependants:			
Children	57%	76%	66%
No-children	43%	24%	34%
Smoking status:			
Smoker	15%	17%	27%
Ex-smoker	-	33%	
Non-smoker	85%	50%	73%
Economic status:			
Employed	60%	49%	56%
Other	40%	51%	44%
N	60	70	4014

Table 3: Question variants and the number of respondents in each

	Round 1 no ranking of lifetime benefit PTO numeraire = 100	Round 2 ranking of lifetime benefit PTO numeraire = 1000
PTO Order 1 5 th vs. 1 st 5 th vs. 3 rd 3 rd vs. 1 st	29	34
PTO Order 2 3 rd vs. 1 st 5 th vs. 1 st 5 th vs. 3 rd	31	36
TOTAL	60	70

Table 4: Ranking results

Age	Round	Rank 1 (%)	Rank 2 (%)	Rank 3 (%)	Rank 4 (%)	Rank 5 (%)
5	1	53	10	23	10	3
	2	39	3	17	24	17
20	1	20	60	10	8	2
	2	14	56	17	13	0
35	1	22	20	53	5	0
	2	43	17	39	0	1
55	1	2	10	10	75	3
	2	3	24	27	58	0
70	1	3	0	3	2	92
	2	1	0	4	13	81

Modal rank is in bold

Table 5: Mean ranks across the two rounds of interviews

Age	Round 1	Round 2
5	2.0	2.8
20	2.1	2.3
35	2.4	2.0
55	3.7	3.2
70	4.8	4.7

Table 6: Reasons for ranking

Reason	Round 1	Round 2
Five years is more valuable to some	0%	21%
Greater benefit to society	8%	17%
Family responsibilities etc.	22%	23%
Lived less life	67%	37%
Other	3%	2%

Table 7: Consistency in PTO responses (%)

$MRS_{5v1} > MRS_{5v3}$	Order 1	Order 2
Round 1	90	97
Round 2	79	72
$MRS_{5v1} > MRS_{3v1}$	Order 1	Order 2
Round 1	72	94
Round 2	76	89

Table 8: The three ages used in the PTO questions

Ages used i.e. 1 st , 3 rd and 5 th ranked	N
5, 35, 70	52
35, 5, 70	19
35, 20, 70	12
20, 55, 70	9
20, 5, 70	7
35, 55, 5	6
5, 20, 70	4
35, 55, 70	3
20, 35, 70	3
20, 55, 5	3
5, 70, 55	2
55,35,70	2
70, 35, 5	2
35,70,5	2
70, 55, 20	1
55, 20, 70	1
35, 20, 5	1
5, 70, 35	1

Figure 1 (a)

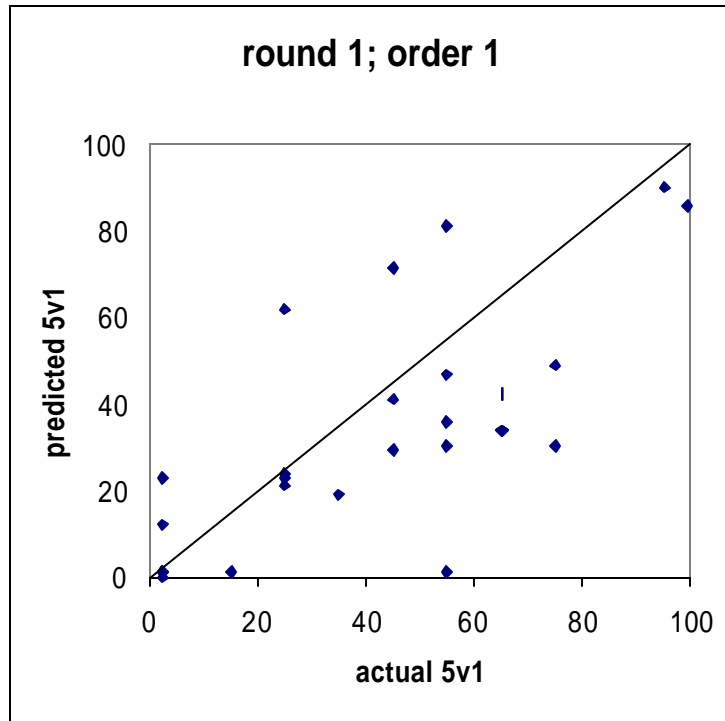


Figure 1 (b)

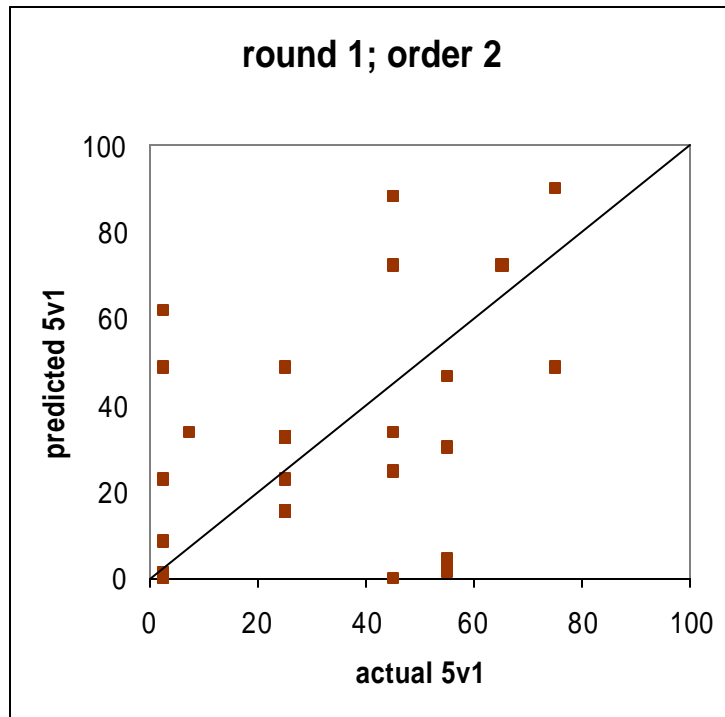


Figure 2 (a)

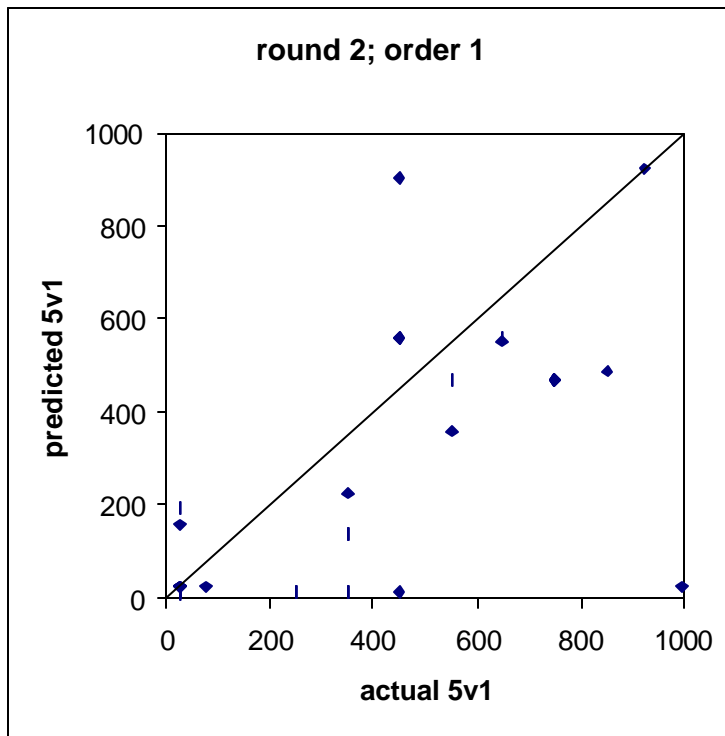
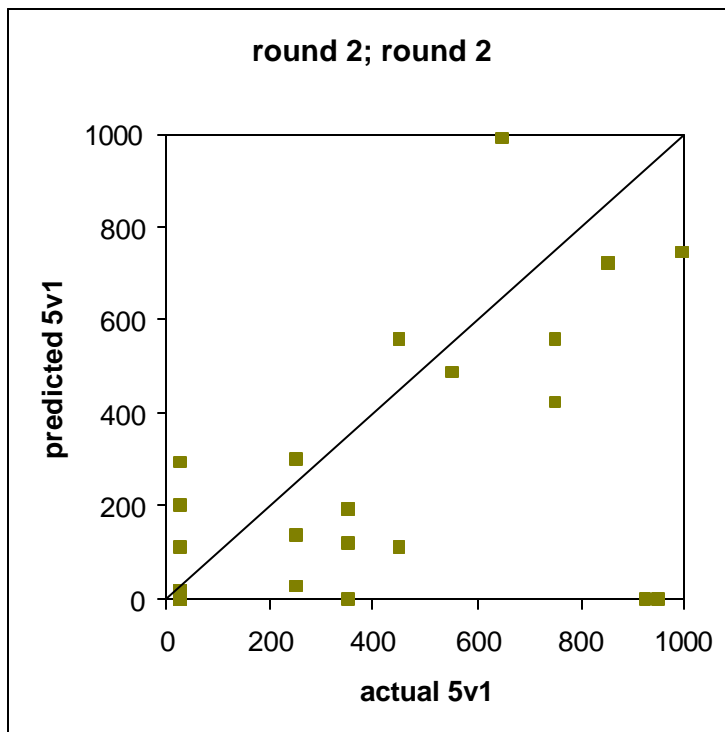


Figure 2 (b)



Appendix: PTO question format

Programme A		Programme B	
Number of ___ year olds who will live for five years		Number of ___ year olds who will live for five years	
	✓ for A	✓ for B	
1000			1000
1000			990
1000			980
1000			970
1000			960
1000			950
1000			900
1000			800
1000			700
1000			600
1000			500
1000			400
1000			300
1000			200
1000			100
1000			50