

# Differences between participants and non-participants in an RCT on physical activity and psychological interventions for older persons

Marieke J.G. van Heuvelen<sup>1</sup>, Jacqueline B.M. Hochstenbach<sup>2,3</sup>, Wiebo H. Brouwer<sup>4</sup>, Mathieu H.G. de Greef<sup>1</sup>, Gertud A.R. Zijlstra<sup>5</sup>, Ellen van Jaarsveld<sup>6</sup>, Gertrudis I.J.M. Kempen<sup>5</sup>, Eric van Sonderen<sup>6</sup>, Johan Ormel<sup>7</sup>, and Theo Mulder<sup>1</sup>

<sup>1</sup>Centre for Human Movement Sciences, University of Groningen, <sup>2</sup>Center for Brain Damage Aftercare, Academic Hospital and University of Groningen, <sup>3</sup>Mesdag Clinic for Forensic Psychiatry, Groningen, <sup>4</sup>Department of Psychology, University of Groningen, <sup>5</sup>Department of Health Care Studies, Medical Sociology section, University of Maastricht, <sup>6</sup>Northern Centre for Health Care Research, University of Groningen, <sup>7</sup>Department of Psychiatry, University of Groningen, The Netherlands

**ABSTRACT. Background and aims:** Volunteer bias in intervention studies on successful aging has been poorly explored. This paper investigated differences between participants and non-participants of the Groningen Intervention Study on Successful Aging (GISSA) over a wide range of demographic, physical, psychological and social subject characteristics. **Methods:** Subjects were recruited among a longitudinal cohort study (Groningen Longitudinal Aging Study) and included 558 men and 711 women, aged 65-96 years, who were invited to participate in the GISSA. Measures were obtained by questionnaires at the moment of invitation and eight years before invitation. Participants were compared with three groups of non-participants: persons who refused to participate, those who did not respond after a reminder, and those who intended to participate but withdrew before pre-test. **Results:** At the moment of invitation, participants were younger, better educated, and functionally and physically more active than the three groups of non-participants. They also had better scores on the physical functioning subscale of the medical outcome scale, better ADL, iADL and vigorous ADL functions and fewer depressive symptoms, and perceived less social support in everyday and problem situations. Participants reported a less strong rate of decline in physical and psychological functioning in the eight years prior to the invitation than did the other groups. **Conclusion:**

*Due to volunteer bias, results of intervention studies on successful aging may have limited generalizability. (Aging Clin Exp Res 2005; 17: 236-245)*

©2005, Editrice Kurtis

## INTRODUCTION

The increasing survival of people to old age forms one of the most dramatic demographic changes of the twentieth century. With advancing age, people are more often confronted with chronic diseases, functional decline, disability and loss of independence (1, 2). The elderly, therefore, rely more heavily on health care facilities (3, 4). Promoting successful aging is a great challenge. Indeed, people who have reached old age successfully are characterized by high levels of functioning in both physical and psychosocial domains (5-7). Several studies have shown that disability can be prevented or reduced by enhancing physical fitness through physical activity (8, 9). In the psychological domain, one often focuses on specific age-related symptoms or disorders like memory impairment (10, 11), depression (12) or dementia (13). Also here, it has been shown that specific interventions help people to cope with these symptoms and to overcome barriers for a socially and physically active lifestyle, thus enhancing the quality of life of older people.

The results of these studies are helpful in developing specific intervention programs for elderly persons. However, one important requirement for applying these results properly is that the studies must provide information regarding

---

Key words: Educational program, elderly, physical activity, psychological intervention, volunteer bias.

Correspondence: Dr. M.J.G. van Heuvelen, Centre for Human Movement Sciences, University of Groningen, P.O. Box 196, 9700 AD Groningen, The Netherlands.

E-mail: M.J.G.van.Heuvelen@ppsw.rug.nl

Received October 10, 2003; accepted in revised form June 4, 2004.

---

their generalizability. Indeed, subjects who volunteer in studies on the promotion of successful aging may not be representative of the target population, but may have selected themselves on the basis of certain characteristics. This may lead to inadequate estimates of effect sizes or to the development of intervention programs that are not suitable for the target population, because the programs are too difficult, too demanding or too long.

Volunteer bias may be investigated by comparing participants' characteristics with the characteristics of persons who refuse to participate or who are not reached by the recruitment strategy. Until now, only a few studies have investigated volunteer bias in the context of intervention studies on the promotion of successful aging. Wagner et al. (14) compared participants with non-participants in a health promotion program for persons over 65 years of age. They found that non-participants had lower self-reported health, a lower level of education and a less active social life, and were more often smokers than were participants. Conversely, in a comparable health promotion program Ives et al. (15) found that participants were more likely to have disease histories and risk factors for disease than non-participants. In a younger population, Alexy (16) investigated differences between participants and non-participants in a workplace wellness center, which had facilities for exercise, and found that non-participants experienced less self-efficacy, and were older and less educated than participants.

Although little information is available on the determinants of volunteer bias in intervention studies on the promotion of successful aging, potential determinants may be derived from studies on participation in survey research, adherence or attendance in physical activity programs, and adherence or attendance in psychological interventions.

In the studies of (non-)participation in survey research, the identified determinants of non-participation included older age (17-19), being female (17, 19), lower level of education (19, 20), lower level of cognitive functioning (20-22), more depressive symptoms (21) and a lower level of activities of daily living (ADL) functioning (21).

Research on physical activity programs showed that, compared with adherers, drop-outs (persons who initially started the program but then withdrew) had less self-efficacy (23-28), more symptoms of depression (25, 29), a lower level of perceived stress (30), fewer perceived benefits (23), less positive outcome expectations (24), less support from family, friends or exercise staff (28, 31), a lower level of physical fitness or physical function (25, 26, 29, 30), more medical problems (25), less prior exercise participation (27, 32), a higher body mass index (30), and were more often smokers (23) and on average younger (29).

In studies on adherence in psychological interventions, lower household income, negative treatment attitudes, greater age, lower education and higher anxiety sensitivity (33), living alone and unemployment (34) and stronger religious beliefs (35) were identified as determinants leading to drop-out.

Against this background, the present study attempts to achieve more insight into the role volunteer bias may play in intervention studies aimed at enhancing activity in the elderly. This was done by critically analyzing the participants of the Groningen Intervention Study on Successful Aging study (GISSA).

GISSA investigates whether physical activity and psychological interventions interact in their influence on disability. Disability was operationalized as the restriction to perform activities of daily living in both physical and psychological domains. In GISSA, participants were randomized over four 18-week intervention programs: physical activity, psychological intervention, a combination of physical activity and psychological intervention, and an educational program as control intervention. Since GISSA participants were recruited from a larger pool of respondents in a longitudinal study (Groningen Longitudinal Aging Study) (GLAS), we were able to investigate differences between participants and non-participants over a wide range of potential predictors. Participants and non-participants were assessed at the moment of invitation. However, since in the context of GLAS all persons had also been measured eight years previously, we had a unique opportunity of investigating differences between participants and non-participants – not only with respect to functioning at a certain moment in time, but also with respect to the potential determinants of (non-)participation that appeared across time.

## METHOD

### *Subjects*

The study sample was a selection of follow-up participants in the GLAS. GLAS is a population-based prospective follow-up study of the determinants of health-related quality of life of older people, with particular emphasis on physical and social disability and well-being (36, 37). The source population of GLAS consisted of individuals aged 57 years or older who live in the north of the Netherlands, either independently or in a home for the elderly. Subjects with severe cognitive impairments were excluded [Mini Mental State Examination score of less than 17 (38)]. Figure 1 shows the selection procedure. In 2001, 3227 persons were registered as participants of GLAS. They received a mailed questionnaire, including repeated measurements of baseline questionnaires, and two screening scales on respectively functional and physical activities. Based on the screening scales, a selection of subjects were invited to participate in GISSA. Those who were classified as (i) very active on both scales or (ii) very active on one scale and moderately active on the other scale, were excluded.

### *Procedure*

Invited subjects received an information brochure with a response form which they were asked to return by

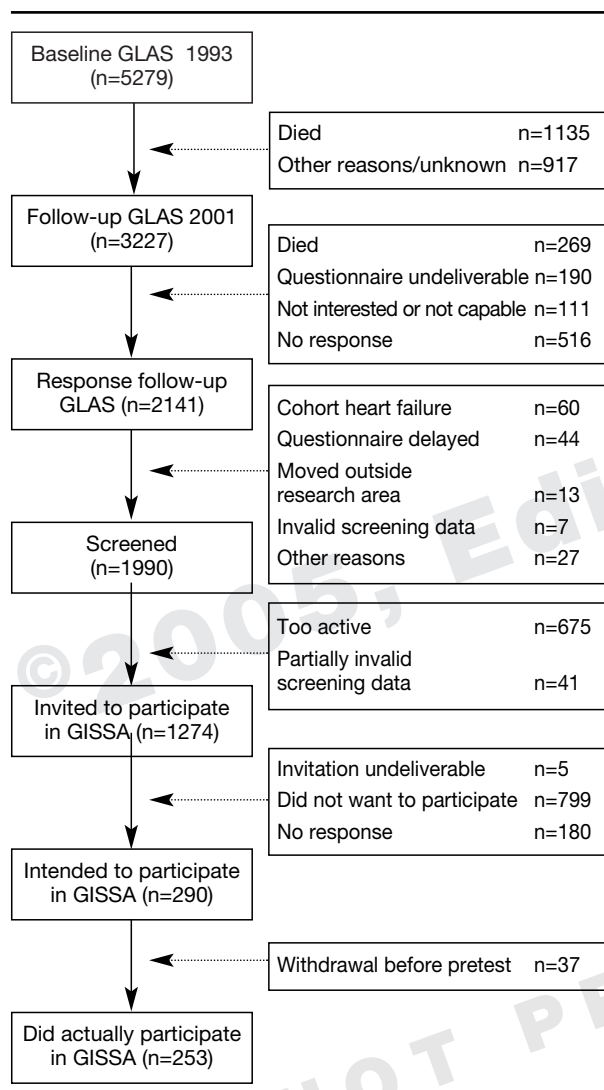


Fig. 1 - Selection procedure.

mail. The brochure included information regarding interventions, test procedures and randomization procedure, and phone numbers to obtain further information from the researchers or advice from an independent physician. A cover letter emphasized the benefits of participating in the program. Non-responders received a single reminder by mail. Before pretest, informed consent was obtained from all participants. The study was approved by the local medical ethics committee.

In the present paper, four groups of subjects who were invited to participate in the GISSA program are compared: subjects who did not respond after the reminder (Non-responders), subjects who did not want to participate (Refusers), subjects who intended to participate

but withdrew before pretest (Withdrawers), and subjects who actually did participate (Participants).

### Measurements

Measurements were obtained at the moment of invitation in GISSA (during follow-up of GLAS) and eight years before invitation (during baseline of GLAS in 1993). Besides demographic, functional and physical activity measures, measurements were obtained in three other domains: physical functioning and health, psychological functioning, and social support.

(i) *Physical functioning and health.* Physical limitations were measured by a subscale of the Medical Outcome Scale (MOS, 39). This subscale contains six items with two possible answers. Sumscores are transformed to a range from 0 (limited on all six items) to 100 (not limited on all six items).

The number of chronic conditions was assessed with a list of 19 conditions (40). For each condition, subjects were asked if they had consulted a family practitioner or specialist or had taken medicine for that condition in the 12 months prior to the assessment. The total number of conditions had a possible range from 0 to 19.

Basic ADLs and instrumental ADLs were assessed with the Groningen Activity Restriction Scale (GARS, 41). The GARS was developed to assess disability in the domains of personal care and domestic activities. It includes eleven basic ADLs and seven instrumental ADLs. Possible scores range from 11-44 (basic ADLs) and from 7-28 (instrumental ADLs). In addition, six items referring to more vigorous activities were also assessed, with four possible answers per item. Possible scores ranged from 6 to 24. For both disability measures, higher scores indicated higher levels of disability.

(ii) *Psychological functioning.* Affective states (feelings of anxiety, depressive symptoms) were assessed with the two subscales of the Hospital Anxiety and Depression Scale (42). On both scales, possible scores range from 0 to 21. General self-efficacy was assessed by means of the self-efficacy scale of Sherer et al. (43), which has a minimum of 16 and a maximum of 80. Mastery or personal control was assessed on Pearlin and Schooler's (44) mastery scale, on which scores may range from 7 to 35. Neuroticism and extroversion were assessed with subscales of the revised version of the Eysenck Personality Questionnaire (45). Both subscales have possible scores from 0 to 12. On all scales for psychological functioning, higher scores indicated higher levels of the concept concerned.

(iii) *Social support.* Social support was assessed using the 12-item Social Support List for Interactions (SSL-12-I) (46). It includes three subscales: everyday support; support in problem situations; and esteem support. For both subscales, the theoretical range varies from 4 to 16. Higher scores indicate more intense social interactions.

### Data analysis

Data were analyzed using SPSS 8.0. Differences in gender, living situation, level of education, functional activity and physical activity between the four groups – non-responders after reminder (Non-responders= NR), those who refused to participate (Refusers= RF), those who intended to participate but withdrew before pretest (Withdrawers=WD), those who actually did participate (Participants=P) – were investigated with Pearson’s chi-square test. Differences in age between the four groups were investigated using analyses of variance with post-hoc tests and Tukey corrections for multiple comparisons. Multivariate and univariate analyses of variance were used to investigate differences between the four groups in physical functioning and health, psychological functioning and social support at the moment of invitation (in which gender was added as a second factor and age was covaried). Additionally, post-hoc tests with Tukey corrections for multiple comparisons were performed. Similarly, mul-

tivariate and univariate analyses of variance with post-hoc comparisons were performed to investigate differences between the groups for measurements eight years before invitation. Lastly, multivariate and univariate analyses of variance were used to investigate the interaction effect of group x time of measurement (moment of invitation versus eight years before invitation) with post-hoc comparisons on the differences between the moment of invitation and eight years before invitation. To avoid “data fishing”, univariate results were only computed for significant multivariate effects, and post-hoc analyses were made only for significant univariate effects. *P*-values of less than 5% were considered to indicate statistical significance.

### RESULTS

Table 1 shows the characteristics of the sample. There were no significant differences in gender or living situation between the four groups. They differed significantly with respect to age, Participants being slightly younger than the

Table 1 - Sample characteristics at moment of invitation.

	Non-responders	Refusers	Withdrawers	Participants	Total	Differences between groups: test results
Number						
Total	180	799	37	253	1269	
Men	75 (42%)	360 (45%)	14 (38%)	109 (43%)	558 (44%)	$\chi^2 = 1.42$
Women	105 (58%)	439 (55%)	23 (62%)	144 (57%)	711 (56%)	$p = 0.702$
Age at invitation (years)						
Mean	74.68	75.07	75.57	73.66	74.75	
Standard deviation	6.80	6.68	5.63	5.66	6.50	$F = 3.21$
Range	65-94	65-96	65-91	65-92	65-96	$p = 0.022$
Living situation (%)						
Alone	42.6	39.3	39.4	39.7	39.8	
With partner	54.0	58.9	60.6	57.2	58.0	$\chi^2 = 3.75$
With other person(s)	3.4	1.8	0	3.1	2.2	$p = 0.711$
Education (%)						
No elementary school	2.2	1.5	2.7	0	1.3	
Elementary	43.9	30.2	21.6	17.0	29.3	
Vocational training	30.0	28.4	40.5	25.7	28.4	
High-school	17.8	28.6	29.7	41.5	29.7	$\chi^2 = 66.9$
(Under)graduate degree	6.2	11.3	5.4	15.8	11.3	$p = 0.000$
Functional activity (%) <sup>a</sup>						
Least active (score ≤15)	62.6	53.1	45.9	38.0	51.1	
Moderately active (15 < score ≤18)	28.1	36.6	48.6	49.2	38.3	$\chi^2 = 30.0$
Very active (score >18)	9.4	10.3	5.4	12.8	10.5	$p = 0.000$
Physical activity (%)						
Inactive (no walking, cycling or sports)	53.1	51.1	51.4	36.0	48.4	
Moderately active (walking/cycling or irregular sports without supervision)	43.4	44.0	45.9	54.8	46.2	
Active (sports under supervision)	3.4	4.8	2.7	9.2	5.4	$\chi^2 = 24.0$
						$p = 0.000$

<sup>a</sup>Minimum score = 5 (inactive), maximum score = 25 (very active).



other groups. *Post-hoc* tests showed only significant differences for age between Participants and Refusers ( $p=0.014$ ). There were significant group differences in level of education, the Participants being more highly educated. For both functional and physical activities, there were significant group differences. Participants were less often classified in the least active group and more often in the most active group.

Table 2 illustrates the group differences in physical functioning, psychological functioning and social support, at the moment of invitation and eight years before invitation. Multivariate results showed a significant main effect of group on the measurements in physical functioning, psychological functioning and social support at the moment of invitation (Pillai's Trace  $F=2.273$ ;  $df=42,3744$ ;  $p=0.000$ ). Table 3 lists the univariate results of the anal-

Table 2 - Differences between four subgroups at moment of invitation ( $t=0$ ) and eight years before invitation ( $t=-8$ ): mean (standard deviation).

	Non-responders	Refusers	Withdrawers	Participants
<b>Physical functioning</b>				
Medical Outcome Scale <sup>a</sup>				
( $t=0$ )	47.1 (35.0)	47.3 (32.9)	34.8 (24.4)	56.6 (30.3)
( $t=-8$ )	72.5 (25.7)	70.3 (26.6)	64.4 (27.8)	74.1 (25.9)
Number chronic conditions <sup>b</sup>				
( $t=0$ )	1.18 (1.4)	1.37 (1.4)	1.57 (1.6)	1.34 (1.4)
( $t=-8$ )	1.00 (1.0)	1.10 (1.1)	1.54 (1.2)	1.00 (1.1)
ADL <sup>b</sup>				
( $t=0$ )	15.0 (5.4)	15.0 (5.9)	15.2 (5.6)	13.4 (3.7)
( $t=-8$ )	12.1 (2.3)	12.4 (3.0)	12.5 (2.2)	12.0 (2.6)
IADL <sup>b</sup>				
( $t=0$ )	12.4 (6.0)	11.9 (5.7)	12.3 (5.3)	10.1 (4.1)
( $t=-8$ )	9.22 (3.6)	9.03 (3.6)	9.71 (3.9)	8.52 (2.8)
Vigorous activities <sup>b</sup>				
( $t=0$ )	10.8 (5.2)	10.5 (4.9)	10.9 (4.2)	8.70 (3.6)
( $t=-8$ )	8.04 (3.1)	8.31 (3.2)	8.10 (2.2)	7.44 (2.5)
<b>Psychological functioning</b>				
Depressive symptoms <sup>c</sup>				
( $t=0$ )	5.40 (4.0)	4.98 (4.0)	6.53 (4.2)	4.81 (3.6)
( $t=-8$ )	4.22 (2.9)	4.01 (3.2)	4.54 (3.3)	4.00 (3.3)
Feelings of anxiety <sup>c</sup>				
( $t=0$ )	4.65 (4.0)	0.43 (3.8)	5.69 (4.3)	4.44 (3.7)
( $t=-8$ )	3.56 (3.2)	3.75 (3.5)	4.00 (3.6)	4.37 (3.8)
General self-efficacy <sup>c</sup>				
( $t=0$ )	54.6 (10.4)	55.4 (11.1)	54.9 (9.8)	56.0 (11.2)
( $t=-8$ )	60.8 (10.6)	61.8 (10.7)	61.7 (11.4)	61.3 (11.3)
Mastery <sup>c</sup>				
( $t=0$ )	21.9 (4.7)	22.6 (4.6)	21.8 (3.6)	22.4 (4.3)
( $t=-8$ )	25.7 (4.8)	25.3 (4.9)	25.5 (4.1)	24.8 (5.0)
Neuroticism <sup>c</sup>				
( $t=0$ )	3.31 (3.0)	3.40 (3.0)	4.00 (3.5)	3.70 (3.1)
( $t=-8$ )	3.31 (2.9)	3.59 (3.0)	3.78 (3.3)	4.00 (3.1)
Extraversion <sup>c</sup>				
( $t=0$ )	6.17 (3.1)	5.77 (3.1)	5.78 (3.3)	5.88 (3.2)
( $t=-8$ )	6.79 (3.2)	6.34 (3.1)	6.31 (3.1)	6.51 (3.1)
<b>Social Support</b>				
Everyday <sup>d</sup>				
( $t=0$ )	9.72 (2.1)	10.0 (2.1)	9.66 (2.0)	9.61 (1.9)
( $t=-8$ )	9.45 (2.0)	9.55 (1.8)	9.86 (1.8)	9.42 (1.8)
Problems <sup>d</sup>				
( $t=0$ )	8.62 (2.8)	8.87 (2.6)	9.06 (2.2)	8.17 (2.2)
( $t=-8$ )	7.21 (2.2)	7.39 (2.0)	7.31 (2.1)	7.09 (1.8)
Esteem <sup>d</sup>				
( $t=0$ )	9.07 (2.5)	9.23 (2.4)	9.14 (2.4)	9.16 (2.3)
( $t=-8$ )	8.55 (1.9)	8.62 (2.0)	8.43 (2.4)	8.55 (1.9)

<sup>a</sup>Higher scores indicate better functioning; <sup>b</sup>Higher scores indicate worse functioning; <sup>c</sup>Higher scores indicate higher level of concept; <sup>d</sup>Higher scores indicate more social interaction.

yses of variance for this moment of measurement. For all measures of physical functioning except for the number of chronic conditions, group differences were significant, Participants being the best functioning group. *Post-hoc* analyses showed that Participants differed significantly from Non-responders and Refusers for all measures of physical functioning, and from Withdrawers on the medical outcome scale and vigorous activities. For psychological functioning, the groups differed significantly on depressive symptoms only. On average, Withdrawers had the most depressive symptoms and Participants the least. Although *post-hoc* analyses showed only non-significant differences between pairs of groups, the difference between Withdrawers and Participants reached borderline significance ( $p=0.054$ ). For both social support in everyday situations and social support with problems, the groups differed significantly, Participants perceiving the least support. *Post-hoc* analyses showed that Participants differed significantly from Refusers on both support in everyday situations and social support with problems.

Multivariate results showed a significant main effect of group on the measurements in physical functioning, psychological functioning and social support eight years before invitation (Pillai's Trace  $F=2.056$ ;  $df=42,3744$ ;  $p=0.000$ ). Table 4 lists the univariate group differences eight years before the invitation. Considering these univariate results, the groups differed significantly with respect to instrumental and vigorous ADLs and the number of chronic diseases at this point in time. Compared with the other groups, Participants had on average the best scores

for both instrumental and vigorous activities of daily living (Table 2). *Post-hoc* analyses showed only significant differences between Participants and Refusers in vigorous activities. Withdrawers were characterized by more chronic conditions than the other groups (Table 2) – a difference that was significant compared with both Participants and Non-responders.

The multivariate results of the analysis of variance showed a significant interaction effect of group x time of measurement (moment of invitation versus eight years earlier) on measurements in physical functioning, psychological functioning and social support (Pillai's Trace  $F=1.476$ ;  $df=42,3744$ ;  $p=0.025$ ). Table 5 lists the univariate results and *post-hoc* comparisons of this interaction effect. All measures of physical functioning except number of chronic conditions revealed significant group x time interaction effects. For all these variables, *post-hoc* analyses showed significant differences between Participants and Non-responders and between Participants and Refusers, Participants showing the smallest rate of decrease in functioning across the eight years preceding the invitation (Table 2). Psychological functioning showed significant group x time interaction effect for feelings of anxiety and mastery. Withdrawers had a larger increase in feelings of anxiety than the other groups (Table 2). According to the *post-hoc* analyses, this was significant in comparison with Participants. Furthermore, Participants and Non-responders differed significantly with respect to change in feelings of anxiety, with less change for Participants (Table 2). For mastery, Non-responders had

Table 3 - Results of analyses of variance at moment of invitation.

	Main effect group <sup>a</sup> ( $df=3,1259$ )		Multiple comparisons with Tukey corrections ( $p$ -values)					
	F	$p$	NR <sup>b</sup> vs RF <sup>b</sup>	NR vs WD <sup>b</sup>	NR vs PC <sup>b</sup>	RF vs WD	RF vs PC	WD vs PC
<b>Physical functioning</b>								
Medical Outcome Scale	7.432	0.000	1.000	0.135	0.010	0.087	0.000	0.000
Number chronic conditions	1.391	0.244						
ADL	4.188	0.006	1.000	0.996	0.014	0.997	0.000	0.232
IADL	6.756	0.000	0.636	1.000	0.000	0.966	0.000	0.095
Vigorous activities	8.753	0.000	0.846	1.000	0.000	0.961	0.000	0.047
<b>Psychological functioning</b>								
Depressive symptoms	2.798	0.039	0.532	0.368	0.383	0.077	0.929	0.054
Feelings of anxiety	1.603	0.187						
General self-efficacy	0.425	0.735						
Mastery	1.718	0.161						
Neuroticism	1.011	0.387						
Extraversion	0.859	0.462						
<b>Social Support</b>								
Everyday	4.100	0.007	0.218	0.998	0.947	0.674	0.017	0.999
Problems	5.273	0.001	0.610	0.756	0.245	0.968	0.001	0.172
Esteem	0.442	0.723						

<sup>a</sup>controlled for age and gender; <sup>b</sup>NR= Non-responders, RF= Refusers, WD= Withdrawers, PC= Participants.

Table 4 - Results of analyses of variance eight years before invitation.

	Main effect group <sup>a</sup> (df=3,1259)		Multiple comparisons with Tukey corrections ( <i>p</i> -values)					
	F	<i>p</i>	NR <sup>b</sup> vs RF <sup>b</sup>	NR vs WD <sup>b</sup>	NR vs PC <sup>b</sup>	RF vs WD	RF vs PC	WD vs PC
<b>Physical functioning</b>								
Medical Outcome Scale	2.010	0.111						
Number chronic conditions	3.063	0.027	0.699	0.037	1.000	0.089	0.601	0.031
ADL	1.469	0.221						
IADL	2.767	0.041	0.899	0.833	0.117	0.597	0.120	0.154
Vigorous activities	5.233	0.001	0.649	0.999	0.125	0.973	0.000	0.532
<b>Psychological functioning</b>								
Depressive symptoms	0.764	0.514						
Feelings of anxiety	2.171	0.090						
General self-efficacy	0.449	0.718						
Mastery	1.363	0.253						
Neuroticism	1.832	0.140						
Extraversion	1.215	0.303						
<b>Social Support</b>								
Everyday	1.019	0.383						
Problems	1.862	0.134						
Esteem	0.594	0.619						

<sup>a</sup>controlled for age and gender; <sup>b</sup>NR= Non-responders, RF= Refusers, WD= Withdrawers, PC= Participants.

Table 5 - Results of analyses of variance of differences between moment of invitation and eight years before invitation.

	Interaction effect of group x time of measurement <sup>a</sup> (df=3,1259)		Multiple comparisons with Tukey corrections ( <i>p</i> -values)					
	F	<i>p</i>	NR <sup>b</sup> vs RF <sup>b</sup>	NR vs WD <sup>b</sup>	NR vs PC <sup>b</sup>	RF vs WD	RF vs PC	WD vs PC
<b>Physical functioning</b>								
Medical Outcome Scale	4.167	0.006	0.735	0.848	0.021	0.514	0.033	0.071
Number chronic conditions	0.860	0.461						
ADL	5.024	0.002	0.935	0.997	0.008	1.000	0.002	0.405
IADL	5.306	0.001	0.823	0.902	0.004	0.989	0.002	0.636
Vigorous activities	6.660	0.000	0.203	1.000	0.000	0.765	0.004	0.097
<b>Psychological functioning</b>								
Depressive symptoms	1.237	0.295						
Feelings of anxiety	4.323	0.005	0.485	0.772	0.016	0.308	0.081	0.043
General self-efficacy	0.781	0.505						
Mastery	3.622	0.013	0.032	1.000	0.017	0.604	0.824	0.417
Neuroticism	0.771	0.510						
Extraversion	0.053	0.984						
<b>Social Support</b>								
Everyday	3.019	0.029	0.481	0.542	0.981	0.143	0.138	0.655
Problems	1.956	0.119						
Esteem	0.107	0.956						

<sup>a</sup>controlled for age and gender; <sup>b</sup>NR= Non-responders, RF= Refusers, WD= Withdrawers, PC= Participants.

the highest decrease in feelings of mastery, which was significantly different compared with both Refusers and Participants. For social support, the group x time interaction effect was only significant for social support in ev-

eryday situations in all groups, except Withdrawers, who showed an increase in support (Table 2). However, post-hoc comparisons did not reveal significant differences between groups.

## DISCUSSION

The purpose of the present paper was to investigate volunteer bias in a RCT study on the promotion of successful aging. Results showed that, at the moment of invitation, Participants were younger, better educated, functionally and physically more active, had better scores on the physical functioning subscale of the MOS, better ADL function, fewer depressive symptoms, and perceived less social support in everyday and problem situations than Non-participants. In the physical domain, these differences were partly present eight years before invitation for instrumental and vigorous ADL. However, it was indicated that, across the eight-year period, Participants showed a slower rate of decline for the physical functioning MOS subscale and ADL functioning than Non-participants. In the psychological domain, a similar same pattern became visible: Participants showed a slower decline in functioning during the eight years prior to invitation.

The post-hoc multiple comparisons showed differences particularly between Participants on one hand and Non-responders and Refusers on the other hand. Withdrawers revealed only a few differences respect to the other groups. One explanation for this may be that this group was slightly smaller in sample size and thus had less power in the multiple comparisons. For instance, the non-significant difference between Participants and Withdrawers in ADL functioning at the moment of invitation ( $p=0.232$ ) had a larger size effect (0.45) but less power (71% using t-test) than the significant differences between Participants and Non-responders (size effect 0.36, power 97%) or between Participants and Refusers (size effect 0.29, power 94%). The descriptive results showed that Withdrawers had a striking pattern, characterized by more chronic conditions, more depressive symptoms, and more feelings of anxiety than the other groups. The persons in this group were all asked for their reasons for withdrawal. Fifteen (41%) referred to their health, especially disease, frailty or expected hospital admission, and eight (22%) mentioned their partner's health as their main reason for withdrawal.

We identified level of education as a determinant of program participation, in accordance with several other studies on participation (14, 16, 19, 20). However, in research on determinants of program adherence, level of education was, as far as we know, only once identified as an important determinant (33). On one hand, this may reflect differences between determinants of program participation and determinants of program adherence. On the other hand, physical functioning and health are often identified as determinants of participation as well as of adherence.

In several studies, self-efficacy was identified as a determinant of adherence in physical activity interventions (23-28). We did not find any differences in self-efficacy between Participants and the three groups of Non-participants. This inconsistency may also reflect differences between determinants of program participation and de-

terminants of program adherence. Another explanation may be that we measured general self-efficacy instead of specific self-efficacy for physical activity or psychological interventions. Measures of specific self-efficacy may give valuable additional information.

Our results showed that Participants perceived less social support than the other groups, which contrasts the findings of Oka et al. (31) and Rhodes et al. (28). This inconsistency too may reflect differences between determinants of program participation and determinants of program adherence. Lower levels of perceived social support may initially stimulate people to participate, because loss of independence may be more threatening for those without adequate social support or because they are especially attracted to the group character of the intervention. However, for long-term adherence, people may really need social support. An alternative explanation is the simple fact that the better functioning participants did not need any social support and consequently did not perceive it as a possibility that was available.

In general, the heterogeneity of the studies we reviewed and our study hampered comparability between them. Studies differed not only with respect to aiming at program participation or program adherence, but also varied with respect to setting. Our study was conducted in a research setting in which participants were randomized over differing types of interventions. Other studies on volunteer bias in intervention studies on the promotion of successful aging (14-16) were conducted in community settings without randomization. Nevertheless, our results are comparable with those of Alexy (16) and Wagner et al. (14). Another difference between the studies is the presence of a question for help – for instance, in studies on adherence to psychological interventions (33-35), or not – for instance, in our study. Furthermore, program characteristics varied among studies: for instance, health promotion versus physical activity or psychological interventions and differences with respect to frequency, intensity and duration. The way participants select themselves may at least partly depend on setting, presence of a question for help, and program characteristics.

The subsample we asked to participate in GISSA had already participated in GLAS, a longitudinal survey study. Several studies have shown that survey research has to cope with volunteer bias regarding younger age (17, 18) and higher levels of functioning in participants compared with Non-participants (19-22). In GLAS, non-response at baseline was higher in the older age groups and slightly higher for women. Among the non-responders, the proportion of patients with malignant neoplasms was higher. This, however, did not apply to heart disease, chronic respiratory disease, or chronic diseases of the musculoskeletal system (47). Consequently, the invited subsample was not totally representative of the general population. If GLAS non-responders could be persuaded to participate in the survey study – for instance, by more ag-



gressive recruitment strategies – differences between participants and non-participants in our intervention study might be larger. Therefore, we may even have underestimated the actual strength of volunteer bias.

In this study, the research sample was recruited among an existing cohort rather than on a totally new recruitment, which is the case in most studies. Our recruitment strategy resulted in the possibility of investigating differences between participants and non-participants extensively. At the same time, this recruitment strategy may also limit our findings. People who are already participating in a longitudinal study may react differently to the request to participate in an intervention study than “new” people. On one hand, they may be disposed to refuse to participate, because they feel that they spend enough time on helping other people do research. On the other hand, they may tend to collaborate earlier because they feel committed. Nevertheless, these considerations are expected to explain only to a small extent why people decide to participate or not, since there may be many other reasons. For instance, our results indicate that functional problems may keep persons from participating, and these impediments will appear in new recruitments as well.

In GLAS, persons with severe cognitive impairment were excluded. This may limit our findings, especially if one is interested in the effects of intervention programs to promote successful aging that are also suitable for cognitively impaired persons. However, this particular population requires specific treatment, and will therefore often be excluded in advance.

In GISSA, we excluded the most active persons, since the interventions were designed for persons at risk of “unsuccessful” aging. The results showed that, among the less active persons, the least active more often did not respond or refused to participate than participants. If we had not excluded the most active persons, the rate of participation would have been higher.

We did not study why subjects decided to participate or not. On one hand, subjects may decide to participate for their own benefit, to become more active, fitter or healthier, or because they expect they will enjoy the activities. On the other hand, their motivation may be “to contribute to scientific research on successful aging”, which may be related to the higher educational level of participants. Also, reasons not to participate may be miscellaneous. Investigation of reasons for (non)participation may add valuable insights which could be used in strategies to enhance participation rates.

The conclusion of this study is that the people who volunteered to participate in GISSA selected themselves on demographic, physical and psychosocial characteristics. This will limit the generalizability of the results regarding the effects of the proposed interventions. In fact, every RCT study is threatened with respect to its external validity. However, if we can identify the characteristics of those who tend to be underrepresented, we can adapt our recruitment

strategies to include more persons with these characteristics and thus enhance generalizability. This may have important consequences for the implementation of intervention programs to promote successful aging in community settings. These programs should reach people who might benefit the most from this kind of interventions. They include not only inactive and less functioning persons, but also persons at risk due to a strong decline in functioning. However, our results showed that these persons might not be reached without adequate recruitment strategies. Especially the latter group, persons now functioning at normal levels but declining relatively fast, needs special attention. They will often be overlooked and, with time, they will be more and more difficult to reach. If persons at risk for “unsuccessful aging” could really be reached - for instance, with intervention by the regular health care authorities - one should bear in mind that the intervention programs developed in RCT research without adequate consideration of external (population) validity, may not be suitable for this specific population.

## ACKNOWLEDGEMENTS

The Groningen Intervention Study on Successful Aging (GISSA) is conducted by the Centre for Human Movement Sciences and the Department of Neuropsychology of the University of Groningen and the Center for Brain Damage Aftercare of the Academic Hospital and University of Groningen. GISSA is financially supported by the Dutch Organisation for Scientific Research (NWO). GLAS is conducted by the Northern Centre for Healthcare Research (NCH) and various Departments of the University of Groningen (RUG), The Netherlands. GLAS and its substudies are financially supported by the Dutch government (through NESTOR), the University of Groningen, the School of Medicine, the Dutch Cancer Foundation (NKB/KWF), and the Dutch Organization for Scientific Research (NWO).

## REFERENCES

1. Jette AM. Disability trends and transitions. In Binstock R, George L, eds. Handbook of aging and the social sciences, 4th edition. San Diego: Academic Press, 1996: 94-116.
2. Goebeler S, Jylha M, Hervonen A. Medical history, cognitive status and mobility at the age of 90. A population-based study in Tampere, Finland. *Aging Clin Exp Res* 2003; 15: 154-61.
3. Rice DP, Laplante MP. Medical expenditures for disability and disabling comorbidity. *Am J Public Health* 1992; 82: 739-41.
4. Brodsky J, Habib J, Hirschfeld M, Siegel B. Care of the frail elderly in developed and developing countries: the experience and the challenges. *Aging Clin Exp Res* 2002; 14: 279-86.
5. Avlund K, Holstein BE, Mortensen EL, Schroll M. Active life in old age, combining measures of functional ability and social participation. *Dan Med Bull* 1999; 46: 345-9.
6. Berkman LF, Seeman TE, Albert M, et al. High, usual and impaired functioning in community-dwelling older men and women: findings from the MacArthur Foundation Research Network on Successful Aging. *J Clin Epidemiol* 1993; 46: 1129-49.
7. Starr JM, Deary IJ, Macintyre S. Associations with successful ageing in the “Healthy Old People in Edinburgh” cohort: being well, fit and healthy. *Aging Clin Exp Res* 2003; 15: 336-42.
8. Binder EF, Schechtman KB, Ehsani AA, et al. Effects of exercise training on frailty in community-dwelling older adults: results of a randomized controlled trial. *J Am Geriatric Soc* 2002; 50: 1921-8.

9. Carlson JE, Ostir GV, Black SA, Markides KS, Rudkin L, Goodwin JS. Disability in older adults 2: Physical activity as prevention. *Behav Med* 1999; 24: 157-68.
10. Foos PW. Effects of memory training on anxiety and performance in older adults. *Educ Gerontol* 1997; 23: 243-52.
11. Schmidt IW, Berg IJ, Deelman BG. Prospective memory training in older adults. *Educ Gerontol* 2001; 27: 455-78.
12. Dick LP, Gallagher-Thompson D, Thompson LW. In Woods R, ed. *Psychological Problems of Ageing. Assessment, treatment and care.* Wiley: Chichester, 1999: 253-92.
13. Woods RT. Psychological "therapies" in dementia. In Woods R, ed. *Psychological problems of ageing. Assessment, treatment and care.* Wiley: Chichester, 1999: 311-44.
14. Wagner EH, Grothaus LC, Hecht JA, LaCroix AZ. Factors associated with participation in a senior health promotion program. *Gerontologist* 1991; 31: 598-602.
15. Ives DG, Traven ND, Kuller LH, Schulz R. Selection bias and non-response to health promotion in older adults. *Epidemiology* 1994; 5: 456-61.
16. Alexy BB. Factors associated with participation or nonparticipation in a workplace wellness center. *Res Nurs Health* 1991; 14: 33-40.
17. Helliwell B, Aylesworth R, McDowell I, Baumgarten M, Sykes E. Correlates of nonparticipation in the Canadian Study of Health and Aging. *Int Psychogeriatr* 2001; 13 (Suppl. 1): 49-56.
18. Herzog AR, Rodgers WL. Age and response rates to interview sample surveys. *J Gerontol* 1988; 43: S200-5.
19. Koval JJ, Ecclestone, NA, Paterson, DH, Brown B, Cunningham DA, Rechnitzer PA. Response rates in a survey of physical capacity among older persons. *J Gerontol* 1992; 47: S140-7.
20. Norton MC, Bretnier JCS, Welsh KA, Wyse BW. Characteristics of nonresponders in a community survey of the elderly. *J Am Geriatr Soc* 1994; 42: 1252-6.
21. Hebert R, Bravo G, Korner-Bitensky N, Voyer L. Refusal and information bias associated with postal questionnaires and face-to-face interviews in very elderly subjects. *J Clin Epidemiol* 1996; 49: 373-81.
22. Launer LJ, Wind AW, Deeg DJH. Nonresponse pattern and bias in a community-based cross-sectional study of cognitive functioning among the elderly. *Am J Epidemiol* 1994; 139: 803-12.
23. Bock BC, Marcus BH, Pinto BM. Maintenance of physical activity following an individualized motivationally tailored intervention. *Ann Behav Med* 2001; 23: 79-87.
24. Deshamais R, Bouillon J, Godin G. Self-efficacy and outcome expectations as determinants of exercise adherence. *Psychol Rep* 1986; 59: 1155-9.
25. Jette AM, Rooks D, Lachman M, et al. Home-based resistance training: predictors of participation and adherence. *Gerontologist* 1998; 38: 412-21.
26. McAuley E. Self-efficacy and the maintenance of exercise participation in older adults. *J Behav Med* 1993; 16: 103-13.
27. Oman RF, King AC. Predicting the adoption and maintenance of exercise participation using self-efficacy and previous exercise participation rates. *Am J Health Promotion* 1998; 12: 154-61.
28. Rhodes RE, Martin AD, Taunton JE. Temporal relationships of self-efficacy and social support as predictors of adherence in a 6-month strength-training program for older women. *Percept Mot Skills* 2001; 93: 693-703.
29. Williams P, Lord SR. Predictors of adherence to a structured exercise program for older women. *Psychol Aging* 1995; 10: 617-24.
30. King AC, Kiernan M, Oman RF, Kraemer HC, Ahn D. Can we identify who will adhere to long-term physical activity? Signal detection methodology as a potential aid to clinical decision making. *Health Psychol* 1997; 16: 380-9.
31. Oka RK, King AC, Young DR. Sources of social support as predictors of exercise adherence in women and men aged 50 to 65 years. *Womens Health* 1995; 1: 161-75.
32. Rejeski WJ, Brawley LR, Ettinger W, Morgan T, Thompson C. Compliance to exercise therapy in older participants with knee osteoarthritis: implications for treating disability. *Med Sci Sports Exerc* 1997; 29: 977-85.
33. Grilo CM, Money R, Barlow DH, et al. Pretreatment patient factors predicting attrition from a multicenter randomized controlled treatment study for panic disorder. *Compr Psychiatry* 1998; 39: 323-32.
34. Tehrani E, Krussel TE, Borg L, Munk-Jorgensen. Dropping out of psychiatric treatment: a prospective study of a first-admission cohort. *Acta Psychiatr Scand* 1996; 94: 266-71.
35. Mosher-Ashley PM. Therapy termination and persistence patterns of elderly clients in a community mental health center. *Gerontologist* 1994; 34: 180-9.
36. Kempen, GJMJ, Ormel J, Brilman EI, Relyveld J. Adaptive responses among Dutch elderly: the impact of eight chronic conditions on health-related quality of life. *Am J Public Health* 1997; 87: 38-44.
37. Ormel J, Kempen GJMJ, Penninx WJH, Brilman EI, Beekman ATF, van Sonderen E. Chronic medical conditions and mental health in older people: disability and psychosocial resources mediate specific mental health effects. *Psychol Med* 1997; 27: 1065-77.
38. Folstein MF, Folstein SE, McHugh, PR. Mini Mental State: a practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12: 189-98.
39. Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey. Reliability and validity in a patient population. *Med Care* 1988; 26: 724-35.
40. van den Berg J, van den Bos GAM. Gezondheidsenquête: het (meten van het) voorkomen van chronische aandoeningen, 1974-1987. Maastricht Gezondheidsstatistiek 1989; 8: 4-21. (in Dutch)
41. Kempen GJMJ, Miedema I, Ormel J, Molenaar W. The assessment of disability with the Groningen Activity Restriction Scale. Conceptual framework and psychometric properties. *Soc Sci Med* 1996; 43: 1601-10.
42. Spinhoven Ph, Ormel J, Sloekers PPA, Kempen GJMJ, Speckens AEM, van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. *Psychol Med* 1997; 27: 363-70.
43. Sherer M, Maddux JE, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers RW. The self-efficacy scale: construction and validation. *Psychol Rep* 1982; 51: 663-71.
44. Pearlin LI, Schooler C. The structure of coping. *J Health Soc Behav* 1978; 19: 2-21.
45. Eysenck SBG, Eysenck HJ, Barrett P. A revised version of the psychoticism scale. *Pers Individ Differ* 1985; 6: 21-9.
46. Kempen GJMJ, van Eijk LM. The psychometric properties of the SSL12-I, a short scale for measuring social support in the elderly. *Soc Indic Res* 1995; 35: 3-312.
47. Kempen GJMJ, Miedema I, van den Bos GAM, Ormel J. Relationships of domain-specific measures of health to perceived overall health among older subjects. *J Clin Epidemiol* 1998; 51: 11-8.