

Measurement of informal care: an empirical study into the valid measurement of time spent on informal caregiving

Bernard van den Berg a,* and Pol Spauwen b

a Centre for Health Economics Research and Evaluation, Faculty of Business, University of Technology, Sydney, Australia

b Department for Prevention and Health Services Research, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

*Correspondence to: Centre for Health Economics Research and Evaluation, Faculty of Business, University of Technology, Sydney, PO Box 123, Broadway, NSW 2007, Sydney, Australia. E-mail: Bernard.vandenBerg@chere.uts.edu.au

Summary

The incorporation of informal care into economic evaluations of health care is troublesome. The debate focuses on the valuation of time spent on informal caregiving, while time measurement, a related and may be even a more important issue, tends to be neglected. Valid time measurement is a necessary condition for the valuation of informal care. In this paper, two methods of time measurement are compared and evaluated: the diary, which is considered the gold standard, and the recall method, which is applied more often. The main objective of this comparison is to explore the validity of the measurement of time spent on providing informal care. In addition, this paper gives empirical evidence regarding the measurement of joint production and the separation between 'normal' housework and additional housework due to the care demands of the care recipients. Finally, the test–retest stability for the recall method is assessed. A total of 199 persons giving informal care to a heterogeneous population of care recipients completed the diary and the recall questionnaire. Corrected for joint production, informal caregivers spent almost 5.8 h a day on providing informal care. If one assumes that respondents take into account joint production when completing the recall questionnaire, the recall method is a valid instrument to measure time spent on providing informal care compared to the diary. Otherwise, the recall method is likely to overestimate the time spent on providing informal care. Moreover, the recall method proves to be unstable over time. This could be due to learning effects from completing a diary.

Copyright # 2005 John Wiley & Sons, Ltd.

'After dinner, he may find himself drinking Brazilian coffee, smoking a Dutch cigar, sipping a French cognac, reading The New York Times, listening to a Brandenburg Concerto and entertaining his Swedish wife – all at the same time, with varying degrees of success.' [1, p. 79]

Introduction

Incorporation of informal care in economic evaluations of health care is troublesome. The debate focuses on the valuation of time spent on informal caregiving, see for instance [2–5]. A perhaps more important and related but often neglected issue is the measurement of time spent on providing informal care. Valid time measurement is a necessary condition for the valuation of informal care. This is also recognised by McDaid [4] and Van den Berg et al. [5].

Evidently, a misestimating of the amount of informal care provided, multiplied by the theoretical correct value per hour results in a miscalculation of the total value of informal care in an economic evaluation.

In this paper, two methods of time measurement will be compared and evaluated: the diary and the recall method. The main objective of this comparison is to discuss the validity of the measurement of time spent on providing informal care. First, we try to validate the recall method compared to the diary. The diary is the gold standard for the measurement of time use [6,7], mainly because it collects time allocation data in a structured way and involves a relatively short recall period. The validation of the recall method is of importance because the diary is more time consuming for respondents than the recall method and therefore less useful for applied research. Even more so, especially if time spent on informal caregiving is just one of the many topics of a survey.

We will also present empirical evidence about the measurement of joint production, which is a persistent problem in the measurement of time in general and in the measurement of informal caregiving. A problem that is specific to the measurement of informal care is the separation between 'normal' housework that somebody does anyway and additional housework that is due to the care demands of the care recipient. We tried to solve this problem by separating 'normal' housework and housework due to informal caregiving in the diary.

Another aim of this paper is to present empirical evidence about the stability over time of the results of the recall method. In other words, we tested the recall method for test–retest reliability. Finally, as a spin off, we will give an indication about the amount of time informal caregivers spent on providing care to a heterogeneous population of care recipients.

The study sample consisted of 199 informal caregivers providing care to a population heterogeneous in terms of disease characteristics and consisting mainly of people with a chronic disease. They completed the diary at one and the recall questionnaire at two moments in time. This implies that we choose for within subject comparisons in our analyses.

The outline of the paper is as follows. In the next section, we describe and discuss the diary and recall method, against a backdrop of available literature. The data are presented and described in the following section. The section following next gives the results of the comparison between the diary and recall method. The results of the recall method's test–retest reliability are described in the penultimate section. In the final section, we discuss the results and present some conclusions.

Methods

Both the diary and recall method are written surveys. We introduce them in this section in more detail, against the backdrop of literature about time measurement.

Time measurement

It is not easy to measure an individual's time use. There are different methods to measure time use (for an overview, see [6–10]). The most important methods are the diary, which is considered the gold standard, and the recall method [6,8]. This gold standard is, however, not universally accepted; an alternative approach is presented by Homan [11, p. 77]. In a diary, respondents are asked to write down all their activities during a specified period of time. The diary has two

important disadvantages; it requires a lot of time and effort from the respondents and it is very costly for researchers. Therefore, measuring time with a diary is not always feasible. This could well be true for the measurement of informal care time, as it may put an extra burden on informal caregivers. A less demanding method like the recall method might be preferred in this context.

The recall method entails respondents being asked how much time they spent on a list of activities during, for example, the previous day or week. Obviously, a major concern with this retrospective way of questioning is its validity because of recall bias. Another concern is the less systematic way of questioning of the recall method compared to the diary.

Another problem in time measurement is how to measure the time spent on tasks that are carried out simultaneously. Doing several activities at the same time or during a specific period of time is called joint production [10]. This is an important issue in informal caregiving. Thus while looking after a person with Alzheimer informal caregivers may perform other (informal care) tasks like cleaning or watching television. Robinson [7, pp. 46–48] suggests that respondents correct for this joint production when completing a recall questionnaire and he presents empirical evidence for this suggestion. In many studies using a recall method, respondents report time use that add up to more than 168 h (7 × 24) a week.

It seems that especially activities that are usually performed in combination with or as secondary to other activities, like watching television, looking after children or resting, contribute to this outcome. Joint production can be measured with a diary. In a diary, respondents may be asked to report all their activities during a certain amount of time, for instance, one quarter of an hour. However, the researcher has to decide how to allocate the time to the activities performed during that period. In practice, most applications of the diary ask respondents only about their main activities instead of all their activities [10].

Another difficulty in measuring time in the context of informal caregiving is that many care recipients receive informal care from different informal caregivers. To get a complete picture, one has to ask all of the informal caregivers to complete a diary or a recall questionnaire. Various instruments have been developed to measure the time spent on informal caregiving, such as the Caregiver Activities Time Survey (CATS) [12], the Caregiver Activity Survey (CAS) [13], and the Resource Utilization in Dementia (RUD) [14]. These instruments are all examples of the recall method. Unfortunately, they have not been validated by comparing them with the diary. Their test–retest reliability has been assessed or the relation between the time spent on informal caregiving and the severity of the care recipient's illness has been studied. All these instruments measure caregiving for people in a homogeneous sample in terms of diseases, e.g. Alzheimer or Dementia.

The informal care diary

We developed an informal care diary to collect valid information about the time informal caregivers spend on providing informal care during a typical 24-h period. A 24-h period was divided into 96 units of 15 min. Three types of informal care tasks were distinguished as follows: (1) support with activities of daily living (ADL), e.g. personal care; (2) support with instrumental activities of daily living (IADL), e.g. managing home adaptations; and (3) housework (HDL), e.g. cleaning the house. We added some general categories of other potential time uses, like sleeping, paid work and unpaid work. To get more precise information about the time spent on informal caregiving, we divided ADL and IADL tasks into four

subcategories each, and HDL tasks into six subcategories. In the diary, the columns contained the tasks and the rows contained the units of time. See Appendix A for a page from the diary.

To account for joint production, respondents could indicate two or more activities in the same quarter of an hour. This was explained in the introduction to the diary. We also gave an example of joint production to prevent respondents from thinking that they were allowed to put only one cross in every quarter of an hour. Owing to this possibility, the total amount of informal care provided could add up to over 24 h a day. We corrected for joint production with the equation:

$$\text{Corrected activity A} = (96 \text{ quarters} / \text{total number of activities per day} * \text{activity A})$$

The total number of activities per day is the sum of the respondents' crosses in all of the quarters of an hour. As a day consists of 96 quarters of an hour, the minimum number of activities per day is also 96. Dividing the time in a diary in units of 15 min is a choice of the researcher. But, the more joint activities an informal caregiver indicates per quarter, the less she or he conforms to the choice of the researcher. This is the intuitive rationale behind Equation (1), which is consistent with the economic law of diminishing marginal productivity. Thus, according to Equation (1), with two crosses in one quarter, each cross counts for 14.8 min. With three crosses in one quarter, they count for 14.7 min each, and so on.

As mentioned before, it is difficult to separate 'normal' HDL tasks from informal care HDL tasks. We tried to solve this problem by splitting the HDL tasks column in normal care tasks, provided in the informal caregiver's own interest, and informal care HDL tasks, performed solely to meet the care demands of the care recipient. The latter column was given a blue colour to stress the difference between the two. This enabled us to compare the recall method with the diary as to the first, the truly informal care and the second, the informal care part of housework added to the 'normal' housework. Thus, we could find out whether or not respondents take into account this difference when completing the recall questionnaire.

To get a representative picture of the informal care provided during a week without making too heavy a demand on the respondents, we asked them to report their time use for two specific days or two reserve days. Moreover, we divided our sample in 21 subgroups covering all possible combinations of 2 days a week. We asked the informal caregivers to fill out the diary six times a day: during breakfast, lunch, and dinner, between these meals, and before going to bed. Thus the diary involves a much shorter recall period compared to the recall method. Another advantage of the diary is that it gives respondents a systematic overview of their time use during an entire day, thereby forcing them to think systematically about their time allocation during that day.

The informal care recall method

The informal care recall method was developed to collect information about the time informal caregivers spent on providing informal care during the week preceding the interview. When completing the recall questionnaire, the respondents in our study were asked emphatically to consider the same week as in the diary. They were asked to consider the same week as a whole rather than the same days for two reasons. First, we suspected that considering the same days would increase the risk of bias due to learning effects. A second, more

practical reason was that it allowed us to compare our results with the results of a previous study using the same recall method.

The recall questionnaire focuses on the same informal care tasks as the diary. Respondents could report their time use in minutes per day or in hours per week, because some activities are perhaps routinely done everyday, while others are less of a routine. The questions are given in Appendix B.

The recall method questions were not presented as the central focus of the survey but as just a part of it, thus to prevent respondents from becoming aware of the research aim.

Diary versus recall method

We will compare the results of the diary and the recall method in two stages. First, in line with Robinson's argument that respondents account for joint production when they complete the recall questionnaire [7], the diary will not be corrected for joint production, while in the second stage the diary will be corrected for joint production. Again, the diary is seen as the gold standard in both cases.

Possible differences between the diary and recall method will be tested for with a t-test [15]. As the same respondents completed the diary and the recall questionnaire for the same week, completing the diary could involve learning effects of completing the recall questionnaire. But because the diary concerned 2 days and the recall method a whole week and because the recall method was embedded in a larger survey on all kinds of informal care aspects, it is not likely that respondents just made a recalculation of the diary when filling out the recall questionnaire. Moreover, in the diary the respondents had to indicate their time use by just putting a cross in a quarter of an hour box as opposed to writing down the exact amount of time per sub-category in the recall questionnaire.

Finally, the respondents had to return the diary before completing the recall questionnaire. The first question of the survey that included the recall questionnaire was whether or not they had already returned the diary. If not, they were asked to do so before going on. They had been sent the diary and survey separately at different times. We checked whether the diary and recall questionnaire were returned separately. If not, they were excluded from the analysis.

Alternative approach

Homan [11, p. 77] suggests a completely different approach. He argues that time should be measured in an aggregated way rather than in the disaggregated ways as discussed above. In other words, he proposes to use just one question about the aggregated time spent on a certain task, instead of many different questions, in order to avoid double counting. According to Homan, another advantage of this approach is that it leaves the classification of the activities performed to the respondents. In case of informal care, some respondents may consider certain tasks as informal care, while others may not. To test Homan's approach, the survey included the following question:

'How much time did you spend on housework during the last week?'

The question focuses on housework instead of informal care to prevent the respondents from becoming aware of our research aim. We test for possible

differences between the answers to this question and the results of the diary and recall method.

Test-retest reliability

We also tested for consistency of the recall method over time. Five months prior to the present study, the respondents had completed the same recall questionnaire as part of another survey. This enables us to test the recall method for test-retest reliability. For the recall method to be a valid method to measure informal care time, a second condition is stability over time. Moreover, a test-retest of the recall method's reliability over time was performed to counter any critique regarding the possible learning effects with a within-sample comparison.

To test the recall method for stability over time, we also asked whether the informal caregivers were of the opinion that the health status of the care recipients had changed since the previous study. Obviously, there is a strong relation between the amount of informal care provided and the care demands relative to the health status of the care recipient. The question about a change in the care recipients' health status, enables us to compare (to identify possible differences) between informal caregivers indicating that the health status of the care recipient had remained the same over time versus caregivers indicating that the health status of the care recipient had improved or deteriorated. Again, possible difference in reported informal care over time will be tested for with a t-test [15].

Background variables

To get a better and broader picture of the informal care situation, we also measured the health-related quality of life of informal caregivers with the EQ-5D and the EQ-VAS [16,17]. We also measured the informal caregivers' subjective burden with a visual analogue scale (VAS) ranging from 0 ('not heavy at all') to 100 ('much too heavy'). Finally, we asked the informal caregivers some general background questions.

Data

Data collection

The data were collected in April 2002. To make sure that our research would enable a comparison between the diary and the recall method instead of just being an exercise in informal caregivers' mathematical skills, we tried to prevent that completing the diary would interfere with completing the recall method. We therefore sent the diary and survey at different times. The diary had to be returned before filling out the survey, as was stressed in the first question of the survey. If the respondents had failed to do so in the survey, both the diary and the survey were excluded from the analysis. As pointed out before, the recall method in the survey covered the same week as the diary.

Study population

Our study population consists of informal caregivers who had participated in an earlier study between October and December 2001, hereafter referred to as the December population. In this study, they had been asked whether they were willing to participate in a future research. Of the 568 informal caregivers who had indicated that they would be willing, 301 caregivers (53%) returned the diary and the survey. Some of them were lost due to the poor quality of their response.

Thus, respondents who failed to indicate their time use during one or more quarters of an hour were excluded. We also lost respondents because they did not return their diary and survey separately and thus circumvented our prevention measure as discussed above. In all, 199 respondents completed the diary and survey correctly. We also conducted a non-response investigation. Reasons for not responding included: the diary was too difficult (11% of the entire population of 568), no time (5%), forgotten (6%), informal caregivers were unable to respond in time due to the pre-specified dates (1%), the care recipient died between the first and the second study (6%), or the informal caregiver was ill (3%). The diary being considered too difficult may indicate a feasibility problem of the method.

Characteristics of the respondents

Table 1 gives the descriptive statistics of the study sample.

Table 1. Background characteristics of informal caregivers and care recipients (n = 199)

Characteristics	Mean
Sex caregivera	70.6
Age caregiverb	57.1
Education caregiverc	
Low	50.9
Medium	24.1
High	25.2
Income of caregiver (in euro)c	
Less than 546	19.9
546–725	15.3
726–900	8.5
901–1135	13.6
1136–1600	13.6
1601–2275	16.5
2276–3000	9.1
More than 3000	3.4
EQ-5D caregiverb	0.77
EQ VAS caregiverb	70.27
Subjective burden VAS caregiverb	54.30
Social relation caregiver and care recipientc	
Care recipient is partner	55.5
Care recipient is parent	21.5
Care recipient is child	12.6
Care recipient is else	10.4
Diseases of care recipientsc–e	
Respiratory diseases	13.1
Circulatory diseases	20.9
Digestive diseases	11.0
Endocrine, metabolic and nutritional diseases	12.6
Musculoskeletal diseases	26.2
Neurological diseases	37.2
Skin diseases	12.0
Psychological diseases	44.5

a Percentage females.

bMean.

c Percentages.

d Reported by the informal caregiver.

eAdd up to over 100% because of co-morbidities.

The majority of the informal caregivers is female with a mean age of 57.1 years (minimum of 21.0 and maximum of 83.0) and care for their partner or their parent. This suggests that our population is quite comparable with other samples of informal caregivers. Moreover, their reported subjective burden is not that high, but 5.6% of the informal caregivers report an EQ-5D score of below 0.3.

The most common diseases of the care recipients as reported by the informal caregivers are psychological, neurological and musculoskeletal diseases. Results diary versus recall method The results of the comparisons of the diary and recall method are presented in this section. First, the results of the diary, uncorrected for joint production, and the recall method are compared.

By implication, the total amount of time spent on informal caregiving could add up to over 24 h a day. Subsequently, we compare the results of the corrected diary with the recall method. In both comparisons only the real informal care part of housework is included.

Then we discuss the comparison between the diary including 'normal' housework and the recall method, to test whether or not respondents take the difference between normal and informal care housework into account when completing the recall method. Finally, we discuss the comparison between answers to the question on aggregated time spent on housework and the results from the diary and recall method.

Uncorrected diary versus recall method

Table 2 presents the results of the uncorrected diary and the recall method. Column four gives their mean differences, and column five the statistical significance of these differences. We found a difference of almost an hour per day in the total time reported to have been spent on informal care between both methods, with the recall method resulting in an hour per day less than the diary. However, this difference is relatively limited, only being 10%, nor is it statistically significant. This suggests that on an aggregated level it is possible to measure informal care with the recall method in a valid way. On the individual and subgroup level, there are differences. For HDL and IADL, the results of the recall method are higher than the results of the diary. But on the individual task levels the diary occasionally scores higher. Although some differences are statistically significant, the subgroup totals are not statistically significant. ADL tasks show the opposite pattern. The diary scores are all higher compared to the recall method and the subgroup total is statistically significant.

Table 2 presents the average results of the whole sample. It should be noted that not all of the tasks are performed by all of the informal caregivers. Some tasks are performed by more informal caregivers, than other tasks. We also checked for differences between the first and the second day of the diary. None of the differences found were statistically significant. So, it seems likely that having filled out the first day does not affect the way respondents fill out the second day. We also checked the diary for difference between the results of a day in the week and a weekend day. There were two statistically significant differences in behavior between days in the week and weekend days. The respondents spent more time on making trips and visiting family or friends on a weekend day than on a day in the week (35.1 versus 15.0 min a day) and they spent no time at all on health care contacts during the weekend ($p=0.0053$ and $p=0.0001$, respectively). The first difference seems to be fact of life and the second one is even more obvious, because one would not expect to visit a general practitioner or physician in the weekends unless it is an emergency.

Table 2. Diary versus recall method; April ($n = 199$)

	Diary (minutes a day)	Recall (minutes a day)	Difference		Correction diary (minutes a day)	Difference	
			Recall – diary	Pr > [t]		Recall – joint production	Pr > [t]
HDL							
Preparation of food and drinks	59.17	61.92	2.75	0.6067	39.63	22.29	0.0000
Cleaning the house	59.36	44.08	-15.28	0.0034	18.30	25.78	0.0000
Washing, ironing and sewing	16.13	24.24	8.11	0.0007	11.21	13.03	0.0000
Taking care of and playing with your own children	18.20	19.76	1.56	0.8262	10.01	9.75	0.0775
Shopping	24.57	31.91	7.33	0.0507	17.88	14.03	0.0001
Maintenance work, odd jobs, gardening	8.63	25.34	16.71	0.0000	5.87	19.46	0.0000
HDL total	186.07	207.25	21.18	0.1833	102.90	104.34	<0.0001
ADL							
Personal care	87.59	47.97	-39.62	0.0042	65.58	-17.61	0.1874
Moving around in the house or going to the toilet	28.61	6.82	-21.79	0.0000	19.11	-12.29	0.0005
Moving or travelling outside the house	21.82	18.67	-3.15	0.4976	15.12	3.56	0.4068
Eating and drinking	66.97	10.60	-56.37	0.0000	44.98	-34.38	0.0000
ADL total	204.99	84.06	-120.93	<0.0001	144.78	-60.72	<0.0001
IADL							
Making trips and visiting family or friends	32.98	20.64	-12.34	0.0186	23.54	-2.90	0.5194
Health care contacts	11.27	0.00	-11.27	0.0000	8.62	-8.62	0.0000
Organising help, aids, house adaptations or taking care of financial matters like insurance	20.46	60.27	39.81	0.4075	14.53	45.75	0.3409
Social support	90.60	118.69	28.08	0.1017	54.54	64.15	0.0001
IADL total	155.31	199.60	44.28	0.3866	101.23	98.37	0.0520
Total	546.37	490.90	-55.47	0.3376	348.91	141.9888	0.0134

Corrected diary versus recall method

We present the results of the diary corrected for joint production in Table 2, column six. Column seven presents the mean difference between the corrected diary and the recall method, while column eight gives the results of the t-test for this difference.

A comparison between the corrected diary with the recall method shows a completely different picture. The differences in case of HDL tasks become larger, while they become smaller in case of ADL tasks. IADL tasks show a mixed pattern.

The two tasks that are more easily combined with other tasks, social support and organising help show even larger differences. Tasks that are less easily combined with other tasks, like making trips and visiting family or friends and health care contacts show smaller differences. Moreover, the recall method overestimates the provision of informal care compared with the corrected diary with more than 2 h/day. This difference is especially due to the HDL and IADL tasks.

Finally, the total difference is statistically significant, which suggests that the measurement of informal care with the recall method compared with the corrected diary overestimates the time spent on caregiving.

Informal care and housework together

Respondents may have difficulties in distinguishing between 'normal' HDL and informal care HDL when they complete the recall questionnaire. This could be an explanation for the overestimation of HDL with the recall method in Table 2. In the diary, informal caregivers indicate that they spent 458.56 min/day on 'normal' HDL. Corrected for joint production, they spent 298.69 min a day on 'normal' HDL. If we add the corrected 'normal' HDL to the informal care HDL, we get a time use of 505.93 min/day. This is over twice as high as the 207.25 min in the recall method.

The difference is also statistically significant ($p=0.0001$). It is therefore unlikely that respondents are unable to make a distinction between 'normal' HDL and informal care when completing the recall method.

One question on HDL

Informal caregivers report that they spent 169.24 min/day on HDL, according to Homan's aggregated approach [11]. This is much lower than the 458.56 min/day ($p=0.0001$) or the 298.69 min/day corrected for joint production as reported in the diary ($p=0.0485$). Because of the statistically significant differences, we may conclude that asking just one aggregated question about an individual's time use seems to result in an underestimation.

Results test-retest reliability

This section presents the results of the test-retest reliability of the recall method. First, we compare the results of all respondents. Subsequently, we compare the results of respondents who indicated that the health status of their care recipient was comparable at the two measurement moments. Test-retest for all respondents Of the 199 respondents, 150 completed the recall questionnaire both times. We analyzed only these results. Table 3 shows that informal caregivers report that between December and April, the time they spent on providing informal care had increased with almost 4 h. Especially, HDL tasks contribute to this difference. For ADL and IADL tasks, the differences are also in favour of the April study, but these differences are smaller and not statistically significant at the 5% level.

Test-retest for respondents caring for care recipients with a stable health status

From the 150 respondents analysed above, five respondents indicated that the health status of their care recipient had improved between the two measurement moments. Seventy-three informal caregivers indicated that the health status of their care recipients had deteriorated during this period. Two respondents failed to answer this question. Seventy respondents indicated that the health status of the care recipients had remained stable. It is therefore likely that the weekly amount of care they provided is comparable at the two moments they completed the recall questionnaire.

Table 4 shows a similar pattern as Table 3. However, the differences are much smaller although still statistically significant in the case of HDL tasks and the total time spent on informal care. In Table 4 IADL tasks are statistically significant as opposed to Table 3. These results suggest that the recall method is not stable over time. In April, informal caregivers reported to spend much more time on providing informal care compared to December. This difference might be explained by the learning effect of completing the diary. Empirical evidence for

this learning effect is that much less respondents failed to complete the recall questionnaire in December compared to April (47 versus two, respectively). This idea is supported by the fact that, in particular, HDL tasks contribute to the differences.

Table 3. Test–retest recall method; December versus April ($n = 150$)

	December	April	Difference	
			April–December	Pr > [<i>t</i>]
HDL				
Preparation of food and drinks	35.72	62.08	26.36	<0.0001
Cleaning the house	23.04	45.32	22.28	<0.0001
Washing, ironing and sewing	9.59	25.50	15.91	<0.0001
Taking care of and playing with your own children	8.25	21.42	13.18	0.0623
Shopping	16.74	35.38	18.64	0.0002
Maintenance work, odd jobs, gardening	6.48	27.56	21.09	<0.0001
HDL total	99.82	217.27	117.45	<0.0001
ADL				
Personal care	17.98	37.43	19.45	<0.0001
Moving around in the house or going to the toilet	6.44	15.39	8.94	0.4619
Moving or travelling outside the house	4.65	3.39	-1.27	0.4862
Eating and drinking	14.28	10.45	-3.84	0.1754
ADL total	43.36	66.65	23.29	0.0939
IADL				
Making trips and visiting family or friends	10.75	19.86	9.11	0.0837
Health care contacts	5.86	21.86	16.00	0.0002
Organising help, aids, house adaptations or taking care of financial matters like insurance	4.23	0.00	-4.23	0.0119
Social support	3.99	75.96	71.97	0.2590
IADL total	24.83	117.69	92.85	0.1494
Total	168.01	401.61	233.60	0.0010

Table 4. Test–retest recall method for care recipients with a stable health status; December versus April ($n = 70$)

	December	April	Difference	
			April–December	Pr > [<i>t</i>]
HDL				
Preparation of food and drinks	40.99	53.78	12.79	0.0705
Cleaning the house	18.21	40.33	22.12	0.0003
Washing, ironing and sewing	7.59	20.39	12.79	0.0001
Taking care of and playing with your own children	9.99	25.21	15.23	0.0712
Shopping	14.96	26.11	11.16	0.0073
Maintenance work, odd jobs, gardening	5.04	21.07	16.04	<0.0001
HDL total	96.77	186.90	90.13	<0.0001
ADL				
Personal care	18.36	34.58	16.22	0.0013
Moving around in the house or going to the toilet	7.78	2.20	-5.57	0.0044
Moving or travelling outside the house	6.14	4.38	-1.77	0.6155
Eating and drinking	20.43	9.18	-11.24	0.0193
ADL total	52.71	50.34	-2.36	0.7980
IADL				
Making trips and visiting family or friends	11.13	22.99	11.86	0.2326
Health care contacts	4.67	19.69	15.02	<0.0001
Organising help, aids, house adaptations or taking care of financial matters like insurance	3.07	0.00	-3.07	0.0008
Social support	2.88	9.18	6.31	0.0035
IADL total	21.75	51.86	30.11	0.0051
Total	171.23	289.10	117.87	<0.0001

Discussion and conclusion

The incorporation of informal care in economic evaluations of health care is troublesome and has been subject of debate for quite some time. Hitherto, the debate has focused mainly on the valuation of informal care and tended to ignore the related issue of the measurement of time spent on providing informal care. This paper tries to fill this gap.

We compared and evaluated two main methods of measuring time spent on providing informal care: the diary and the recall method. The main objective was to explore the validity of the measurement of time. We therefore compared the recall method with the diary, the gold standard for the measurement of time use.

In all, 199 people giving informal care to a heterogeneous population completed a diary and recall questionnaire. The diary is considered the gold standard for measuring the time spent on providing informal care. The recall method is also a valid tool to measure time spent on informal care if it is assumed that respondents take into account joint production when completing the recall questionnaire. Otherwise, the recall method will overestimate the time spent on providing informal care. This is a problem that needs to be addressed, if informal care is to be incorporated in economic evaluations.

In the context of informal caregiving, the separation between 'normal' housework and additional housework due to the care demands of the care recipients is often neglected and requires specific attention. It is likely that respondents are unable to separate between normal HDL and informal care when they complete the recall questionnaire. We did not find any evidence that this distinction is not clear to respondents when they complete the diary.

The gold standard for time measurement is not universally accepted. An alternative approach is to use just one aggregated question to measure time spent on a certain category of tasks. We tested for this approach. It appears to result in an underestimation of time use.

Finally, we assessed the test–retest stability of the recall method. The recall method proves to be unstable over time. This could be due to learning effects from completing the diary. But, the interpretation of the results of studies applying the recall method needs caution.

A weakness of the present study is that there may be a feasibility problem with the diary. More than 10% of the sample indicated not to participate in our research because they found the diary too difficult to complete. Another point of debate is the correction for joint production. This correction, despite being based on the economic idea of decreasing marginal productivity, is a choice of the researcher. Future research may provide empirical evidence regarding this correction or develop other ideas about how to correct for or deal with joint production. A final issue is that care recipients may have access to more than one informal caregiver. The empirical study in December 2001 asked the care recipients about their social relations including primary and secondary informal caregivers. The rationale behind this question was that we expected the care recipients to know best what care they get from whom. We were aware that this might not be the case, especially when the primary informal caregiver organises the care for the care recipient.

The care recipients were not interviewed again in April 2002. Therefore, we do not have information about possible changes in the number of caregivers, which could bias our re-test results. The application of the recall method seems to

involve an overestimation of the total amount of informal care provided. Future research with other designs, like between subject comparisons, could provide more evidence about the validity of the recall method. Another promising area for future research is the optimal number of tasks a diary and recall questionnaire should contain in order to get the most valid outcomes. Future work may also consider the merits of looking at a matched control group of non-carer households and record their use of time on household tasks as an alternative or complement for asking caregivers to divide time between informal care and 'normal' household activities. The results could serve as a basis for identifying incremental activities related to informal caregiving.

From a research practice point of view, it might not always be possible to use the diary in economic evaluations, because completing the diary is very time consuming for respondents and the method is relatively expensive for researchers. To illustrate the latter, we used student assistants to enter the data in a data base. It took them approximately 1 h to enter the data of two diaries, while they entered eight complete surveys in 1 h. The recall questionnaire was just one-fifth of the total survey. So, importing a diary takes 20 times as much time compared with importing the recall method.

Whether or not the benefits of increasing precision of using the diary instead of the recall method outweigh the additional costs is debatable and depends on the research objectives and the outcome of future research regarding the validity of the recall method.

Future research should provide more evidence regarding the validity of the recall method. The recall questionnaire may be a valid method to measure informal care if informal caregivers first fill out a diary. It might be helpful if future research involves evidence of stable overestimations of the recall method compared with the diary (for certain care tasks or sub-populations) resulting in corresponding correction factors.

Acknowledgements

A previous draft of this paper was presented at the Fourth International Conference of the international Health Economics Association (iHEA) 2003 in San Francisco and at a seminar of the Department of Health Policy and Management (iBMG), Erasmus Medical Center, Rotterdam. We thank Marc Koopmanschap, Wien Limburg, our former colleagues of the Department for Health Services Research (CZO), and two anonymous referees for useful comments on an earlier draft of this paper. We also thank the Department for Prevention and Health Services Research (PZO) of the National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands, for their funding.

Notes

- a. Simultaneously does not necessarily mean at exactly at the same moment. It can also mean during the same period of time (for instance 15 min).
- b. Of course, many other changes may have occurred in the mean time. Care recipient may, for example, get more or less or different informal care or more or less professional care.

References

1. Linder SB. *The Harried Leisure Class*. Columbia University Press: New York, 1970.
2. Smith K, Wright K. Informal care and economic appraisal: a discussion of possible methodological approaches. *Health Econ* 1994; 3(3): 137–148.
3. Posnett J, Jan S. Indirect cost in economic evaluation: the opportunity cost of unpaid inputs. *Health Econ* 1996; 5(1): 13–23.
4. McDaid D. Estimating the costs of informal care for people with Alzheimer's disease: methodological and practical challenges. *Int J Geriatr Psychiatry* 2001; 16: 400–405.
5. Van den Berg B, Brouwer WBF, Koopmanschap MA. Economic valuation of informal care: an overview of methods and applications. *Eur J Health Econ* 2004; 5(1): 36–45.
6. Juster FT, Stafford FP. The allocation of time: empirical findings, behavioral models, and problems of measurement. *J Econ Lit* 1991; XXIX: 471–522.
7. Robinson JP. The validity and reliability of diaries versus alternative time use measures. In *Time, Goods, and Well-Being*, Juster FT, Stafford FP (eds). Institute for Social Research The University of Michigan: Ann Arbor, 1985.
8. Juster FT. Conceptual and methodological issues involved in the measurement of time use. In *Time, Goods, and Well-Being*, Juster FT, Stafford FP (eds). Institute for Social Research The University of Michigan: Ann Arbor, 1985.
9. Gronau R. Home production – a survey. In *Handbook of Labor Economics*, Ashenfelter O, Layard R (eds). Elsevier Science Publishers: Amsterdam, 1986.
10. Kooreman P, Wunderink S. *The Economics of Household Behaviour*. MacMillan Press: Houndmills, 1996.
11. Homan ME. *The Allocation of Time and Money in One-Earner and Two Earner Families; An Economic Analysis*. Erasmus University Rotterdam: Rotterdam, 1988.
12. Clipp EC, Moore MJ. Caregiver time use: an outcome measure in clinical trial research on Alzheimer's disease. *Clin Pharmacol Ther* 1995; 58(2): 228–236.
13. Davis KL, Marin DB, Kane R et al. The Caregiver Activity Survey (CAS): development and validation of a new measure for caregivers of persons with Alzheimer's disease. *Int J Geriatr Psychiatry* 1997; 12(10): 978–988.
14. Wimo A, von Strauss E, Nordberg G et al. Time spent on informal and formal care giving for persons with dementia in Sweden. *Health Policy* 2002; 61(3): 255–268.
15. Rice JA. *Mathematical Statistics of Data Analysis*. Duxbury Press: Belmont, 1995.
16. EuroQol Group. A new facility for the measurement of health-related quality of life. *Health Policy* 1990; 16(3): 199–208.
17. Essink-Bot ML, Stouthard ME, Bonsel GJ. Generalizability of valuations on health states collected with the EuroQol-questionnaire. *Health Econ* 1993; 2(3): 237–246.