

## S1. Supplementary file

### S1.1. Data imputation

Missing questionnaire data for people who continued to participate at each wave were multiply imputed by chained equations. Wave 1 equations for imputing participant questionnaires data included baseline characteristics (socio-economic status, education, living status, income, tenure, age, sex, living status, diagnostic subtype, and carer status), service use (at the item level) and costs (at the category level). Carers' questionnaire imputation equations included socio-economic status of the carer, carer status, lost work and employment status variables. Imputation models of data from Waves 2 and 3 differed depending on the source of the data (participant or carer questionnaire). Imputation of service use data at item level was carried out for Wave 1 data from participant and carer questionnaires, and similarly data from carer questionnaires at Waves 2 and 3 were imputed at item level. Data from the participant questionnaires were mostly (apart from medications) provided by the relatively small number of participants without participating carers, limiting the number of parameters that could be included in the imputation equations, so that cost data from Waves 2 and 3 of participant questionnaires were imputed at the category level.

Imputation equations at each Wave included predictors from that wave and the dependent variable from the previous waves. Because of changes to questionnaires in Waves 2 and 3 (see Measures section in the main text), imputations of most items at Wave 3 were limited to the sample with the same source at Wave 2. As dyads had been interviewed together about paid and unpaid care (in participant questionnaires) at Wave 1, Wave 2 equations for service use/cost variables from either participant or carer datasets included the dependent variable (that is, service use at item level or cost category level), from the Wave 1 participant data. Equations for Wave 3 included the dependent variable from Wave 1 and from Wave 2 if the same source had completed the question (e.g. the Wave 3 variable would be predicted by the Wave 2 variable only if a carer had completed the Wave 2 and Wave 3 questionnaire). Also, a few questions differed between questionnaires

(provision of unpaid care; medications). Therefore for a small number of cases at Wave 3 (21 from the carer data; 21 cases from the participant data), data were not imputed and so could contain missing values. There were a few instances of extremely low use (by less than 1% of cases) in the observed data where frequency of use could not be imputed; Wave 1 nursing home visits, psychology visits in the Wave 2 carer dataset, and incontinence laundry services in Waves 2 and 3 carer datasets. For these variables, use in the cases missing data was assumed to be zero. Also Wave 3, where use of a fourth outpatient speciality was reported in only 2% of cases, mean imputation was used to complete the cases that were missing frequency of use data when use had been reported.

## S1.2. Inverse probability weights

The inverse probability weighting approach (IPW) was used to account for attrition across follow-up waves. In IPW, the inverse of the probability of being a complete case is used to weight complete cases, and as a result some cases will take more weight than others in the analysis [1]. In this study, a case was considered 'complete' if the participant with dementia completed all or some of the study questionnaire at that wave. Reasons for non-response at follow-up could be related to the outcomes of interest in this study. For instance, severe cognitive disability might impede participation in the study and also be associated with increased care costs. The IPW model assumed that data were missing at random, in other words that the probability of response was conditional on observed variables. For example, assuming cognition was a predictor of non-response, cases with lower cognitive scores would receive a higher weight in the analysis, and vice versa

Candidates for predictors of drop-out at Wave 2 and 3 were determined by univariate logistic regression models, assessing the size and strength of effect on drop-out of baseline participant characteristics (including age, sex and marital status, educational attainment, carer status, area level deprivation, socioeconomic status, dementia sub-type, country) and measures of cognition, wellbeing, income, functional status, living status, tenure, service use, for the previous wave. Cases

where drop-out due to death occurred at the subsequent Wave were excluded from the models. Multivariate logistic regressions of drop-out included variables significant at the 5% level in univariate analyses: age, sex and marital status, educational attainment, carer status, country (England, Scotland or Wales), living status, cognition, function, wellbeing, comorbidities, use of practice nursing, outpatient services and prescription medications. Cases where the participant had died before that wave were assigned a zero weight. Wave-specific weights (Wave 2: mean 0.96, SD 0.565, range 0 - 4.745; Wave 3: mean 0.753, SD 0.668, range 0 - 4.106) were applied in descriptive statistics of costs at each wave.

### S1.3. Model selection

Initial modelling fitted generalised linear covariate models to weekly costs (3 months costs were scaled by 13 to aid computational tractability without affecting interpretability of the results) at each wave. At each wave, cost distributions exhibited departures from normality, skewing to the right (skewness: Wave 1: 6.35, Wave 2: 9.53, Wave 3: 3.79). A modified Park test [2, 3] was applied to determine relationships between the conditional mean and variance of the cost data at each wave. GSEM model fit was assessed applying the Bayes Information Criterion (BIC) with a lower BIC indicating better model fit [4]. The modified Park test gave the gamma distributional family for specifying the mean-variance relationship of Wave 1 and Wave 2 cost data, but the test did not identify a suitable family at Wave 3. GSEM were fitted to the costs data. Of unconditional models with (i) Gaussian family and identity link (ii) lognormal family and log link and (iii) gamma family and log link, the latter had the lowest BIC statistics (Table S1.1) and so the model with gamma-family and log-link was taken forward.

### S1.4. References

- [1] Seaman SR, White IR (2013) Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res* **22**, 278-295.
- [2] Glick H (2007) *Economic evaluation in clinical trials*, Oxford University Press, Oxford ; New York.

- [3] Manning WG, Mullahy J (2001) Estimating log models: to transform or not to transform? *J Health Econ* **20**, 461-494.
- [4] Rabe-Hesketh S, Skrondal A (2012) *Multilevel and longitudinal modeling using Stata. Volume I, Continuous responses.*

## S1.5. Supplementary tables

Table S1.1. Model fit: BIC statistics

Family, link	Original data	Imputed data
Gaussian, identity	51070.1	68183.15*
Lognormal, log	41307.5	55752.44
Gamma, log	40602.5	55136.42

Notes: Results averaged using multiply imputed data (40 complete datasets). Inverse probability weights applied.

\*averaged over 39 imputations, as the model did not converge in one dataset.

Table S1.2 Completion of paid and unpaid care questions by participants and caregivers: response transitions between waves

Completion at Waves 1 and 3	Completion at Wave 2				Total N	Percentages	
	Dyad did not participate*	Both dyad members participated†	Only Participant participated‡	Only Caregiver participated†		Wave 1	Row %
<b>Wave 1</b>	N (row %)	N (row %)	N (row %)	N (row %)	Wave 1	Row %	Column %
Both dyad members participated§	261 (20.4%)	941 (73.7%)	51 (4.0%)	24 (1.9%)	1277	100.0%	83.1%
Only Participant participated‡	77 (29.6%)	18 (6.9%)	165 (63.5%)	0 (0.0%)	260	100.0%	16.9%
<b>Total (row %)</b>	<b>338 (22.0%)</b>	<b>959 (62.4%)</b>	<b>216 (14.1%)</b>	<b>24 (1.6%)</b>	<b>1537</b>	<b>100.0%</b>	<b>100.0%</b>
<b>Wave 3</b>	N (row %)	N (row %)	N (row %)	N (row %)	Wave 3	Row %	Column %
Dyad did not participate*	325 (51.8%)	223 (35.6%)	70 (11.2%)	9 (1.4%)	627	100.0%	40.8%
Both dyad members participated†	8 (1.2%)	672 (97.1%)	10 (1.4%)	2 (0.3%)	692	100.0%	45.0%
Only Participant participated‡	2 (1.3%)	18 (11.7%)	134 (87.0%)	0 (0.0%)	154	100.0%	10.0%
Only Caregiver participated†	3 (4.7%)	46 (71.9%)	2 (3.1%)	13 (20.3%)	64	100.0%	4.2%
<b>Total (row %)</b>	<b>338 (22.0%)</b>	<b>959 (62.4%)</b>	<b>216 (14.1%)</b>	<b>24 (1.6%)</b>	<b>1537</b>	<b>100.0%</b>	<b>100.0%</b>

Note: 8 dyads that did not participate at Wave 1 were excluded from the analyses and are excluded from the table counts.\* Questionnaires were completed by neither the participant nor the caregiver.

† Caregivers completed questions within the Caregiver questionnaire on participant's use of paid and unpaid care.

‡ Participants completed questions within the Participant questionnaires on use of paid and unpaid care on their own.

§ Dyads jointly completed questions on use of paid care and unpaid care within the Participant questionnaire.

Table S1.3. Mean unweighted costs (£, 2014-15) of care during the prior three months over Waves 2 to 3

Cost categories (£)	Wave 2			Wave 3		
	Source	Mean (95% CI)	N	Source	Mean (95% CI)	N
<i>Sub-total and total costs</i>						
<i>Health and social care</i>						
Primary and community health	P,C	150 (137,164)	1199	P,C	155 (131,178)	892
Community mental health	P,C	36 (28,44)	1199	P,C	31 (24,38)	894,899
Community social care*	P,C	208 (177,240)	1199	P,C	258 (217,298)	898
Day care services	P,C	175 (146,205)	1199	P,C	221 (183,260)	902
Hospital services	P,C	387 (281,493)	1199	P,C	373 (267,478)	895,898
Care home stays	P,C	163 (93,233)	1199	P,C	436 (294,577)	897
Total medication†	P,C	58 (50,65)	1199	P,C	54 (46,62)	909
Equipment (Social services & NHS)‡	P,C	16 (14,19)	1199	P,C	18 (15,21)	899
Total services & medications§	P,C	1194 (1048,1339)	1199	P,C	1523 (1323,1723)	876,882
<i>Out-of-pocket costs to the person, relatives &amp; friends</i>						
Equipment (self or family)	P,C	36 (33,40)	1199	P,C	38 (34,42)	899
Condition-related travel participant & caregivers¶	P,C	6 (2,9)	1199	P,C	8 (3,14)	899
Total out-of-pocket	P,C	42 (37,47)	1199	P,C	47 (40,54)	895
<i>Costs of unpaid care and lost working time</i>						
Unpaid care from main caregiver#	C	3940 (3658,4222)	983	C	4613 (4274,4952)	751
Unpaid care from friends/relatives**	C	256 (201,310)	983	C	257 (198,316)	753
Unpaid care, all caregivers ††	C	4195 (3902, 4488)	983	C	4877 (4528,5225)	750

Care from any friends & relatives§§	P	1042 (627, 1458)	216	P	792 (416,1168)	148
Lost work time (caregivers) ¶¶¶	C	102 (69,134)	1194	C	78 (47,109)	908
Lost work time (friends/relatives) ###	C	35 (20,50)	983	C	27 (19,36)	754
<i>Total costs</i>						
Total, participants without a caregiver***	P	2045 (1557,2532)	216	P	1544 (1125,1964)	142
Total, proxy-reported†††	C	5482 (5132,5833)	983	C	6613 (6193,7032)	731,737

Notes: Results of multiply imputed data (40 complete datasets). N reports inverse-probability weighted observations from each complete dataset – where observations differed between complete datasets, the range of observations is reported. P=Participant with dementia; C=Caregiver.

\* Includes costs of respite stays and permanent residence.

† Costs of dementia and CNS medications.

‡ Costs over prior 3 months.

§ Assumes all community care costs fall to social services.

¶¶ Costs of travel to appointments related to problems with thinking, memory and behaviour by participant and caregiver or participant-only if no caregiver was involved.

# Costs of hours of unpaid care by unpaid caregiver. Costs valued at national minimum wage; hours estimated from Wave 1 Participant questionnaires completed in interviews of complete dyads and from Waves 2 and 3 Caregiver questionnaires.

\*\* Costs of hours of unpaid care by other friends and relatives. Costs valued at national minimum wage; hours estimated from Wave 1 Participant questionnaires completed in interviews of dyads of participants and caregivers and from Waves 2 and 3 Caregiver questionnaires.

†† Costs of hours of unpaid care by unpaid caregiver and by other friends and relatives. Costs valued at national minimum wage; hours estimated from Wave 1 Participant questionnaires completed in interviews of complete dyads and from Waves 2 and 3 Caregiver questionnaires. Mean costs of hours of unpaid care estimated for 260 participants without caregivers at Wave 1: £979 (95% CI £633, £1325).

§§ Costs of hours of unpaid care by any friends and relatives. Costs valued at national minimum wage; costs estimated from interviews with participants without a caregiver in the study completing Waves 2 and 3 Participant questionnaires.

¶¶¶ Costs of lost working time by unpaid caregiver valued at national average wage based on Wave 1 participant questionnaires and Waves 2 and 3 Caregiver questionnaires.

### Costs of lost working time by other friends and relatives valued at national average wage based on Wave 1 participant questionnaires and Waves 2 and 3 Caregiver questionnaires.



\*\*\* Costs estimated from interviews with participants without a caregiver in the study completing Participant questionnaires. Excludes costs of lost working time to avoid double-counting with costs of time spent in unpaid care.

††† Costs estimated from Wave 1 Participant questionnaires completed in interviews of dyads of participants and caregivers and from Waves 2 and 3 Caregiver questionnaires. Excludes costs of lost working time to avoid double-counting with costs of time spent in unpaid care.

Table S1.4. Marginal means (95% confidence intervals) (£, 2014-15) from conditional latent growth curve model of total paid service costs

Variable	Wave 1 Mean (95% CI)	Wave 2 Mean (95% CI)	Wave 3 Mean (95% CI)
Sample	72 (65, 79)	96 (88,105)	156 (128, 183)
<i>Sex</i>			
Male	77 (68, 86)	101 (89, 112)	158 (126, 190)
Female	65 (57, 73)	91 (80, 102)	153 (120, 186)
<i>Household status</i>			
Lives with others	67 (59, 74)	90 (80, 100)	146 (116, 176)
Lives alone	90 (70, 110)	118 (92, 143)	185 (123, 247)
<i>Age</i>			
At mean age of the sample	71 (64, 78)	94 (86, 103)	149 (124, 175)
<i>Diagnosis</i>			
AD	59 (53, 65)	79 (71, 86)	125 (102, 148)
VaD	66 (51, 80)	85 (66, 103)	131 (87, 176)
Mixed	88 (71, 104)	105 (88, 122)	150 (108, 193)
FTD	69 (41, 96)	123 (79, 167)	265 (130, 399)
PDD	167 (110, 224)	268 (185, 351)	513 (217, 808)
DLB	87 (59, 115)	177 (113, 242)	433 (154, 711)
Unspecified/other	123 (64, 181)	187 (102, 272)	341 (145, 537)
<i>Caregiver relationship</i>			
Spouse/partner	68 (60, 76)	88 (77, 98)	134 (106, 162)
Family/friend	95 (75, 115)	154 (120, 188)	295 (192, 399)
No caregiver involved	62 (48, 76)	70 (56, 85)	94 (60, 128)

Notes: Results using multiply imputed data (40 complete datasets). Inverse probability weights applied.

Abbreviations: AD=Alzheimer's disease; VaD= vascular dementia; FTD= frontotemporal dementia; PDD Parkinson's disease dementia; DLB dementia with Lewy bodies; Other= Unspecified/other.

Table S1.5. Marginal effects for the difference between sub-groups

Comparison	Mean difference	Lower 95% CL	Upper 95% CL	p-value
Wave 1 Female vs. Male	-12	-23	-1	F(1,26944)=4.515,p=0.034
Wave 1 Lives alone vs. Lives with others	23	1	45	F(1,5571)=4.348,p=0.037
Wave 1 Mixed vs. AD	28	12	45	F(1,4715)=11.364,p=0.001
Wave 1 PD vs. AD	108	51	164	F(1,71586)=14.052,p=0.000
Wave 1 PD vs. VaD	101	43	159	F(1,51637)=11.769,p=0.001
Wave 1 PD vs. Mixed	79	22	137	F(1,99478)=7.298,p=0.007
Wave 1 PD vs. FTD	98	36	160	F(1,74101)=9.702,p=0.002
Wave 1 DLB vs. PD	-80	-142	-18	F(1,46462)=6.435,p=0.011
Wave 1 Unspecified/other vs.AD	63	5	122	F(1,4645)=4.515,p=0.034
Wave 1 3 Family/friend vs. Spouse/partner	27	4	50	F(1,6361)=5.187,p=0.023
Wave 1 No caregiver involved vs. Family/friend	-33	-53	-12	F(1,7820)=9.568,p=0.002
Wave 2 Mixed vs. AD	26	9	44	F(1,32191)=8.863,p=0.003
Wave 2 FTD vs. AD	45	1	89	F(1,40588)=3.957,p=0.047
Wave 2 PD vs. AD	189	107	271	F(1,19388)=20.189,p=0.000
Wave 2 PD vs. VaD	183	100	266	F(1,17757)=18.538,p=0.000
Wave 2 PD vs. Mixed	163	81	245	F(1,19805)=15.050,p=0.000
Wave 2 PD vs. FTD	144	54	235	F(1,23661)=9.786,p=0.002
Wave 2 DLB vs. AD	98	34	163	F(1,3208)=9.038,p=0.003
Wave 2 DLB vs. VaD	92	27	158	F(1,2712)=7.566,p=0.006
Wave 2 DLB vs. Mixed	72	8	137	F(1,3122)=4.798,p=0.029
Wave 2 Unspecified/other vs.AD	108	23	193	F(1,17744)=6.268,p=0.012
Wave 2 Unspecified/other vs.VaD	102	16	188	F(1,18153)=5.435,p=0.020
Wave 2 Family/friend vs. Spouse/partner	66	28	104	F(1,10581)=11.867,p=0.001
Wave 2 No caregiver involved vs. Family/friend	-83	-117	-50	F(1,7348)=23.862,p=0.000
Wave 3 FTD vs. AD	140	9	271	F(1,74330)=4.364,p=0.037
Wave 3 PD vs. AD	388	96	680	F(1,18137)=6.769,p=0.009
Wave 3 PD vs. VaD	382	89	675	F(1,15267)=6.527,p=0.011
Wave 3 PD vs. Mixed	362	72	653	F(1,20555)=5.979,p=0.014

<b>Comparison</b>	<b>Mean difference</b>	<b>Lower 95% CL</b>	<b>Upper 95% CL</b>	<b>p-value</b>
Wave 3 DLB vs. AD	308	32	583	F(1,3403)=4.787,p=0.029
Wave 3 DLB vs. VaD	301	25	577	F(1,3263)=4.585,p=0.032
Wave 3 Unspecified/other vs. AD	216	24	409	F(1,9220)=4.867,p=0.027
Wave 3 Unspecified/other vs. VaD	210	16	404	F(1,10088)=4.524,p=0.033
Wave 3 Family/friend vs. Spouse/partner	161	54	269	F(1,6481)=8.642,p=0.003
Wave 3 No caregiver involved vs. Family/friend	-201	-297	-105	F(1,4625)=16.777,p=0.000

Abbreviations: CL=confidence limit; AD=Alzheimer's disease; VaD= vascular dementia; FTD= frontotemporal dementia; PDD Parkinson's disease dementia; DLB dementia with Lewy bodies; Other= Unspecified/other