Long-Term Care Across Europe and the U.S.: The Role of Informal and Formal Care

Daniel Barczyk
McGill University and CIREQ

Matthias Kredler
Universidad Carlos III de Madrid

August 17, 2019

Abstract

Large cross-country variation in long-term-care (LTC) policy in conjunction with household-level data on caregiving provides a valuable laboratory for policy analysis. However, comprehensive cross-country data on how care is provided are lacking. In order to close this gap, we draw on data from the Survey of Health, Ageing, and Retirement in Europe (SHARE) and the Health and Retirement Study (HRS) in the U.S. Since care hours are missing for some care forms (especially for nursing-home residents), we propose a selection model to impute these. The model allows selection into care forms to differ by country. Our estimates imply that nursing-home residents have higher care needs, even when conditioning on observed characteristics. In contrast to the bulk of the literature, we also take into account care provision from persons in the same household, which we find to contribute one-third of all care hours. Informal-care provision in Europe follows a steep North-South gradient, the U.S. falling in between Central-European and Southern-European countries. The results are robust to alternative imputation schemes.

Contact information: Daniel Barczyk (daniel.barczyk@mcgill.ca). McGill University, Department of Economics, Leacock Building, Room 321b, 855 Sherbrooke Street West, Montreal, QC, H3A 2T7. Matthias Kredler (matthias.kredler@uc3m.es). Universidad Carlos III de Madrid, Department of Economics, C. Madrid, 126, 28903 Getafe.

I. Introduction

The provision of long-term care (LTC) to frail elderly people is a pressing policy issue. Some countries, such as Sweden and the Netherlands, already spend around 4% of GDP on LTC, and the trend is rising. There have been LTC policy reforms in several countries, e.g. Germany and Spain, and there is a lively debate in many others on how the problem of increasing care needs should be addressed. The rising female labor-force participation may well shift Southern-European countries, which have so far relied heavily on family-provided care, towards care in nursing homes. This will only intensify the debate on if and how formal caregiving should be subsidized by governments.

In general, care can be provided in different settings: informally by family members or friends (*informal care*, IC) or formally. Formal care is again sub-divided in care given by paid helpers in the elderly's home (*formal home care*, FHC) and care given in nursing homes (NH). There is evidence that countries differ widely in how much they make use of different care forms. However, it is hard to come by a quantification on precisely by *how much* countries differ in making use of these various forms of care (see the literature review below). We suspect that progress along these lines has been hampered since data on the intensity of care (e.g. hours of care) is limited. The goal of this paper is therefore twofold. First, we propose a methodology of how to overcome several challenges inherent in current data, the Survey of Health, Ageing, and Retirement in Europe (SHARE) and the Health and Retirement Study (HRS) for the U.S. Next, we use our data to provide a big-picture view of how LTC is provided across Europe and the U.S, in all of its forms. Since our approach encompasses *all* forms of informal care, our results are valuable not only for the literature on LTC but also for the literature on home production in general since we provide hours data for an important component of home production across countries. We also provide estimates by how much gross domestic product would rise if informal care were accounted for.

We combine data from SHARE with the HRS. The two data sets have a very similar structure, but are not identical. We take into account all forms of care: care received in nursing homes, all forms of formal home care (including meals-on-wheels and private domestic help), informal care from household members, and informal care from persons residing outside the household. We document that there is substantial under-sampling of nursing-home residents in SHARE. We correct for this bias by constructing sampling weights that bring SHARE in line with government-provided OECD data on institutional care. We find no evidence of under-sampling of nursing-home residents in the HRS, where the fraction is in line with that provided by the OECD.

There are several other challenges that we have to overcome in order to be able to quantify care arrangements across countries and to make the data in SHARE and HRS comparable. Most significantly, SHARE does not ask for hours of care provided when it comes to care provided by household members. However, care by household members is absolutely essential in order to

obtain a complete picture of care provision. Caregivers often co-reside with those needing attention around the clock. A second issue, inherent to both data sets, is the absence of hours of care for nursing-home residents.

In order to be able to compare the relative importance of various sources of care, we impute for nursing-home residents community-equivalent hours of care, that is, hours of care that they would have received were they observed in the community. A challenge we face here is that our observed sample of community residents is likely subject to the following selection bias: Those who remain in the community are presumably in better shape than those who choose to enter a nursing-home (even when controlling for observed components of dependency). In this case, using the community sample to impute equivalent-community hours of care for nursing-home residents leads to an underestimation. Our first contribution is to estimate a selection model for community residents that takes into account this bias. Our identifying assumption is that for a fixed set of functional limitations, total hours of care have the same mean in all countries (although observed hours in the community are allowed to –and in fact do– vary due to selection). We estimate this model using the semi-parametric approach proposed by Dahl (2002) on the HRS, in which we have complete hours data for all forms of informal care. Using this selection model allows us to impute community-equivalent care hours for nursing-home residents for the U.S. and Europe as well as total community hours in SHARE which we then subsequently use to to back out care hours provided by household members in SHARE.

There are several lesser challenges that need to be addressed, such as, the feature that SHARE provides data for some care forms only on the household level, where we bring these data down to the individual level by exploiting the information on dependency of household members.

Having dealt with the challenges of constructing comparable data and correcting for selection, we turn to the second central contribution of our paper, which is to use the data for a broad view of how care in all of its forms is provided both in terms of frequencies and intensities. Our main findings are the following. We document a strong North-South gradient in the utilization of informal care. In *Northern* countries with high public spending on LTC (Sweden, Denmark, Netherlands, Belgium) informal care accounts for only 22% of all care hours. This number rises to 43% in *Middle* countries (Germany, France, Austria), which exhibit medium-sized public spending on LTC, the *Southern* European countries (Spain, Italy), in which the government spends least on LTC, have the highest percentage of informal-care hours: 81%. The U.S., also a low public spender on LTC, falls between Middle and South with an informal-care share of 54%.[1]

Using our assembled data set, we also document the following additional facts. The North-

[1]See Figure 4 for the data we used to calculate these percentages; Table 4 provides an overview of public LTC programs and motivates the grouping of countries based on their commonalities into the broader regions, *North, Middle* and *South*.

South gradient in IC becomes even stronger when restricting the sample to singles: IC contributes only 8% of all hours of care to singles in the North, 29% in the Middle, 40% in the U.S., and a whopping 77% in the South.[2] The lion's share of informal care hours is provided by spouses for coupled individuals and for singles by children, who tend to be female and of working age.[3] Informal care and nursing-home care are the most important forms of care.[4] Formal home care (FHC) plays a minor role; even in the Middle, which has the highest FHC prevalence, only 12% of all care hours are from FHC. This is despite the fact that a large number of care recipients obtain at least some care from FHC sources (up to half of the sample in North and Middle). The reason is that (i) FHC is often used by less-dependent individuals and (ii) FHC tends to be combined with informal care, FHC contributing often a very low share of hours.

In terms of broad time trends, we find that the fraction of individuals in IC has dropped by about 20 percentage points in the North and the Middle in less than 10 years. In the South, we find a 10-percentage-point decrease in informal care over the same time period. The U.S. however, experienced a slight increase in informal care.

Although we find a strong correlation of government-provided LTC and formal-care uptake across countries, we do not provide causal evidence on this link.[5] However, our results are certainly of great interest for the policy debate since any model of care choice must be consistent with cross-country facts on policy and care choices. In order to further inform this debate, we provide an overview of the institutional setting across the countries in our sample.

We now turn to a review of the literature. A comprehensive overview of how care in all its forms is provided across countries is clearly important for policy makers and economists wanting to study the nexus between policy, economic variables, and care provision. However, to the best of our knowledge, there is no such comprehensive overview so far. We will organize our discussion of the related literature around Table 1. The table presents a non-exhaustive list of articles with cross-country evidence on LTC, showing which sources of care these studies have taken into account. The upshot is that currently no single study has accounted for all forms of care neither in frequencies (case counts) nor in intensities.

First, there are studies drawing on government-provided data (OECD/EU, 2013 and the European Commission, Lipszyc et al., 2012). These provide excellent aggregate data on formal-care

[2]See Figure 8 for the precise hours data.

[3]See Figure 9 for hours for couples and singles and Table 8 for characteristics of children who are the main caregivers.

[4]The fraction of individuals receiving informal care decreases strongly in dependency. In the U.S. this effect is strongest: The proportion of individuals who receives informal care as the primary source of care drops from 90% for the lowest dependency level down to 22% for the highest level.

[5]We note that the causality may go either way: Policies certainly influence LTC choices at the individual level, but populations with a preference for a certain care form (e.g. formal care) may also elect politicians who implement policies favoring this form of care.

provision across countries, both for nursing homes and care at home. However, these data sets lack aggregate data on informal care. As mentioned before, we use the data provided by the OECD on the number of nursing-home residents to correct for the under-sampling of nursing-home residents in SHARE that we document at the country level.

TABLE 1: Sources of care studied

Paper	Data source	IC from someone:		FC at:		IC provided by:	
		inside HH	outside HH	home	NH	child	spouse
Our paper (2019)	SHARE W1-6, HRS, gov'ts	✓ (32.5%)	✓ (18.9%)	✓ (10.1%)	✓ (38.5%)	✓	✓
OECD (2013, 2015)	Governments			✓	✓		
Eur. Comm. (2012)	Governments			✓	✓		
Bolin, Lindgren & Lundborg (2008)	SHARE W1		✓	✓		✓	
Bonsang (2009)	SHARE W1		✓	✓		✓	
Balia & Brau (2014)	SHARE W1		✓	✓		✓	
Crespo & Mira (2014)	SHARE W1,W2	✓*	✓			✓	

This table is a non-exhaustive but representative list of articles that provide cross-country evidence on LTC and take into account various forms of care. *Informal care (IC) from someone inside the household (HH)* means informal care provided from someone living in the same household as the elderly in need of care. *IC from someone outside HH* means informal care provided from someone living in a different household than the care recipient. *Formal care (FC) at home* is care provided by a formal (paid) caregiver at the home of the elderly. *FC at NH* is care in a nursing home. In brackets in second row: percentage of hours in different care forms based on our calculations using waves one and two of SHARE. *In Crespo & Mira (2014)'s sample, co-residency is extremely rare and there is no imputation of hours for co-residing caregivers. They rely on the daily-help filter in the survey design and infer from it that intensive care is provided.

Second, there is a literature on informal caregiving that draws on SHARE's Waves 1 and 2. However, this literature does not consider informal care provided inside the household and care provided by spouses, both of which we find to be very important sources of care. It also restricts itself to formal care at home and omits nursing-home care entirely. Qualitatively these omissions likely do not matter for their results, but quantitatively probably they do. For example, Bolin et al. (2008*a*), Bonsang (2009), and Balia & Brau (2014) study whether informal care from children and formal home care are substitutes or complements, and whether the institutional framework matters for this relationship. They find that informal care at home and formal care at home are rather substitutes than complements[6], which is in line with our result that care tends to be concentrated on one source. If taking informal care provided by a spouse and nursing-home care into account,

[6]The only exception to this finding is Bonsang (2009) finding nursing care to be a weak complement to informal care.

their findings would presumably be strengthened: The spouse is a plausible substitute for formal home care; formal care in a nursing home is a clear substitute for intensive informal care. The results from this literature are in line with our finding that mixing of formal and informal care is rather limited.[7]

A third strand of the literature uses SHARE to study how labor-market outcomes of children who provide informal care are affected and how this trade-off differs across countries. Again, these studies all leave aside spousal care and nursing-home care. Bolin et al. (2008*b*) find that informal caregiving comes with large costs in terms of lost employment opportunities especially in central Europe; in contrast, Crespo & Mira (2014) do not find such an effect in central European countries but rather in southern European countries. In the United States the literature has found that informal caregiving does come with adverse consequences on labor-market outcomes.

Finally, Table 1 shows that none of the studies considers caregiving by spouses. One reason for this is data availability: SHARE does not ask for hours of caregiving by spouses. Another reason is that spouses are usually already retired when their partner becomes dependent so that they do not face the tough trade-off between caregiving and market work that the children are typically confronted with. We provide imputations for care hours from spouses that can be used to address this problem. Similarly, caregiving in nursing homes has not been considered in most studies, most likely because not enough time has passed to make SHARE representative of the nursing-home population. Again, the re-weighting scheme that we propose (using the OECD data) may provide an avenue to remedy this issue.

The article is structured as follows. Section II. provides a description of the data sets, explains our methodology and states the assumptions used. Section III. gives a brief overview of the institutional features of government LTC provision across countries in our sample. In Section IV., we present our results, which is followed by a discussion in Section 5. Section VI. concludes.

II. Methodology

A comprehensive overview of LTC needs to account for informal and formal-care arrangements both at home and in LTC institutions. In order to provide such an overview, we have to overcome several challenges. In the following we first provide an overview of HRS and SHARE. We then lay out the most significant issues that need to be addressed and describe briefly how we tackle each of them. The appendix contains the details.[8]

[7]The substitutability between informal and formal care has also been found in the United States by, for example, Van Houtven and Norton (2004) and Charles and Sevak (2005).

[8]Codes for the procedures described in this section and the resulting data tables are available on request from the authors and will be made publicly available upon publication.

II.1. HRS and SHARE

The Health and Retirement Study (HRS) is a panel study of bi-yearly frequency conducted by the University of Michigan. It began in 1992 and became representative of the U.S. population of ages 50 and above as of 1998. A cohort initially sampled consists of only non-institutionalized individuals and thus excludes the nursing-home (NH) population. However, respondents who subsequently move to nursing homes are retained in the study and interviewed whenever possible, often through a proxy respondent. In order to ensure the representativeness of the nursing-home sample the HRS provides sampling weights, which are currently available for the years 2000-2012.[9] Our sample consists of data collected for the years 2000-2012 (six surveys).

The HRS obtains detailed information about care obtained from others due to functional limitations with regards to activities of daily living (ADL: dressing, bathing, going to bed, eating, walking across a room) and instrumental activities of daily living (IADL: shop for groceries, prepare meals, take medication, manage money, use phone). Respondents are asked about the helpers with the various (I)ADL limitations declared and how many hours each helper provides.[10] An important caveat is that no data on hours of care are collected for nursing-home residents.

The Survey of Health, Ageing, and Retirement in Europe (SHARE) is a panel of bi-yearly frequency of the population aged 50 and above. SHARE was modeled after the Health and Retirement Study (HRS) in the U.S., which makes a comparison between the two possible. SHARE was introduced in 2004 and has currently the following waves available: Wave 1 (2004/05), Wave 2 (2006/07), Wave 3 (2008/09), Wave 4 (2010/11), Wave 5 (2013), and Wave 6 (2015). Wave 3 (SHARELIFE) is special in that it is retrospective: It focuses on respondents' life histories. As of 2015, SHARE covers 27 European countries and Israel. We restrict attention to European countries that are in SHARE since Wave 1, which are: Austria, Belgium, Switzerland, Germany, Denmark, Spain, France, Greece, Italy, Israel, Netherlands, and Sweden. We drop Greece and Switzerland. Greece did not conduct surveys in Waves 4 and 5 and some authors report problems with data collection there (see Balia & Brau, 2014 and Bolin et al., 2008*a*). In Switzerland, there were problems in the data collection on caregiving, which is also noted by Bolin et al., 2008; Bonsang, 2009; and others.

[9] Below we will see that the HRS seems to do a very good job in tracking NH residents since the fraction of the nursing-home population in the HRS is almost identical to OECD data.

[10] Whenever possible we use the cleaned-up data versions provided by the RAND Corporation. Note that the RAND family file does not include information on helpers other than children. Thus, to find out about the importance of, for example, the spouse in providing informal care we use the original HRS data.

TABLE 2: Nursing-home population as of 2015

Country	NH sampled by SHARE (HRS)?	% in NH over 65: in OECD	% in NH over 65: in SHARE (HRS)	Ratio OECD-SHARE (HRS) weighted	Ratio OECD-SHARE (HRS) unweighted
Sweden	yes	4.5	2.6	1.8	2.5
Netherlands	yes	5.3	4.2	1.5	2.0
Denmark	yes	3.9	2.9	1.3	1.5
Belgium	no	8.8	4.3	2.0	2.0
Austria	no	–	2.2	–	–
Germany	yes	4.1	1.6	2.6	3.3
France	no	4.2	2.3	1.9	1.6
Spain	yes	1.8	1.3	1.3	1.2
Italy	no	–	1.1	–	–
U.S.	yes	3.3	2.9	1.1	0.8

Column *NH sampled by survey?*: says if nursing-home residents were eligible in SHARE's baseline and refreshment samples according to Ch. 5 of Börsch-Supan & Jürges, 2005 and Ch. 8 of Malter & Börsch-Supan, 2013. *% in NH over 65*: Proportion of nursing-home residents among population that is above age 65 from: (i) *OECD* statistics from year 2015 (2014 for the Netherlands, Denmark, and Belgium for which 2015 was not available), (ii) our own calculations in *SHARE* using Wave 6 using the calibrated cross-sectional household weights. *Ratio OECD-SHARE*: Gives (i) ratio of two preceding columns (*weighted*) and (ii) the same ratio when not using household weights. The last row shows data for the U.S. from 2011 (no OECD data available after 2011); the *SHARE* columns for the U.S. are from our own calculations using the 2012 HRS RAND files, using household weights where appropriate.

II.2. Non-representativeness of nursing-home population in SHARE

Both SHARE and HRS aim to be representative of the entire population aged 50 and above. Consequently, both surveys make an effort to follow up on respondents moving to nursing homes. However, our findings indicate that SHARE is not (yet) representative of the NH population. This is obviously a concern when trying to establish the relative importance of nursing homes for care provision. Table 2 shows that SHARE's fraction of the NH population is about one-half of that in the OECD data, which draws on aggregate data reported by governments, whereas the HRS is almost spot-on. Furthermore, Table 10 in the appendix shows NH fractions of SHARE and the HRS over time to check whether these converge to aggregate data reported by the OECD. Convergence would be suggestive that the surveys successfully retain respondents who move to a NH. Indeed, in the initial waves of SHARE the fraction of NH residents increases as SHARE follows up with them, but this fraction does not converge to that reported by the OECD.

We also find additional evidence consistent with the notion that under-sampling in SHARE is due to lower retention rates for NH residents. In SHARE, NH residents are substantially more likely to attrite (36%) than community residents (25%). This difference persists when controlling for dependency, age, and other characteristics.[11] To correct for the under-sampling of NH residents,

[11] Sampling problems, such as these, are pervasive in surveys with elderly subjects due to *gatekeeping* problems, see Kars et al. (2016). Gatekeeping refers to the fact that health-care professionals or nursing-home staff prevent eligible subjects from being sampled in studies.

we adjust the SHARE-provided sampling weights upward so that the implied proportion of elderly above 65 in NH matches the OECD numbers for each country in each wave.

Obviously, capturing merely the right fraction of the NH population is insufficient to claim that the results we obtain are representative of this group. The assumption needed for representativeness is the following:

Assumption 1 *Nursing home residents are under-sampled as calculated in OECD-SHARE ratio, but nursing-home respondents are otherwise randomly selected from the nursing-home population in each country.*

This assumption implies that conditional on residing in a NH, the probability of attriting is independent of other respondent characteristics. We cannot reject this condition in our data: When we regress an attrition dummy on a set of observables (age, gender, a disability index –defined below–, education etc.) of NH residents, none of the coefficients is significant.[12]

Quantitatively, our adjustment matters a great deal on how important NH care is. For example, when using our adjusted weights, average daily hours of care that an over-65-year-old in SHARE obtains from NH is 0.42, while it is only 0.15 when using the original SHARE weights – a drop of 64%.

II.3. Determining care needs when hours of care are missing

Beyond SHARE's non-representativeness of the NH population, several further care-related variables present challenges that need to be addressed before a credible comparative analysis of all care forms can be undertaken. At the heart of the problem is that SHARE (and also the HRS, in the case of NH residents) often provides only an indicator variable on whether a certain type of care is utilized, instead of hours of care. In the absence of care hours, however, the importance of different sources of care cannot be adequately assessed if we want to take into account the *intensity* of care. Table 3 provides a broad overview for which types of care data are available across the waves in SHARE. Note that SHARE differentiates between informal care (IC) from *outside* the household (OIC), e.g. from adult children living elsewhere, and informal care from *inside* the household (IIC), e.g. from the spouse or co-residing children, a distinction the HRS does not make.

As in the HRS, SHARE records the nursing-home status of an individual, but not hours of care received. Next, hours regarding formal home care (FHC) and OIC are available only in Waves 1 and 2. An additional complication with OIC is that hours are reported on the household

[12]It is harder to test if there is non-random attrition at the point where individuals leave the community and *enter* a nursing home. The problem is that for a community resident who attrites we do not observe whether she enters a NH in the moment she disappears from the sample or leaves the sample for other reasons.

level instead of the individual level.[13] Hours provided by IIC are lacking in all SHARE waves – this poses a challenge because co-residing caregivers often give the most intensive care (see, for example, Barczyk & Kredler, 2018).[14]

TABLE 3: Overview of care data in SHARE

Type of care	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Level
nursing home (NHC)	dummy	dummy	dummy	dummy	dummy	dummy	indiv.
formal home care (FHC)	hours	hours	–	–	dummy	dummy	indiv.
informal from outside hh. (OIC)	hours	hours	–	freq.	freq.	freq.	hh.
informal from inside hh. (IIC)	dummy	dummy	–	dummy	dummy	dummy	indiv.

Table shows which information on care received is available in SHARE for different forms of care (rows) in the different waves (columns). *dummy*: Question(s) available *if* this form of care was received. *freq.*: Question available if this type of care was received daily, weekly, monthly, or less often. *hours*: Question(s) available on hours of care received of this form of care. *Remark:* For IIC, the survey question only asks for care that was given on daily or almost daily basis; for nursing-home care, we only consider survey information on permanent nursing-home residency (but not short-term stays for rehabilitative purposes). Last column (*Level*) shows if information for a care form is available on the individual level (*indiv.*) or on the household level (*hh*). All information is taken from the SHARE questionnaires.

Broadly speaking, our strategy to close the missing-hours gap is the following. We first assign OIC hours in SHARE's couples to individuals using the dependency information we have on the individuals in the household, which is unproblematic in the large majority of cases. We then estimate a selection model that jointly predicts: (i) the selection of individuals into different forms of care and (ii) the hours of care they receive when residing in the community. We use this model to impute the expected care need of individuals for whom we do not observe hours.

II.3.a. Split hours inside SHARE couples who are both dependent

SHARE usually asks a respondent if she/he (as an individual) has received a certain kind of care; this is the case for NH, FHC, and IIC. However, SHARE does not follow this rule when it comes to OIC (informal care received from persons who reside *outside* the household). For OIC, SHARE asks the respondent about care that all members of the respondent's household received together. Since our analysis is at the individual level, we have to make assumptions on how to assign the care received to one (or both) of the partners in a respondent household with a couple. To do this, we use the information on both partners' limitations with ADLs and IADLs. In more than three-fourths of the cases, this assignment is unproblematic since only one of the partners has limitations. In these cases, we assign all care hours to the partner with limitations. If, however, both partners

[13] In the HRS, care variables are always on the individual level.

[14] Presumably due to this data structure, there is so far very little research in SHARE on care from this important group of caregivers.

report limitations, we assume that both partners received OIC and split the care hours received proportionally using a care-needs index, which we explain now.

In order to obtain a measure of how important each (I)ADL is in determining care needs, we regress the total hours of care on 11 (I)ADL dummies and the dementia dummy, omitting the constant term. The sample of the regression consists of HRS and SHARE community residents for whom the total hours of care are known on the individual level (for SHARE, we can only take singles and couples who answered separately to the question at this point). The coefficients, which turn out to be all positive, are assigned as the weights to the (I)ADLs and the dementia dummy. OIC hours of care at the household level (in SHARE's Waves 1 and 2) are then divided among household members proportionally to the sum of weights over (I)ADLs that each household member has. We now formally state our assumption:

Assumption 2 *Informal care hours from outside the household (OIC) are distributed among couples proportionally to a disability score that equals the predicted values from a regression of care hours on (I)ADL dummies and a dementia dummy (see last paragraph).*

In order to check for robustness of this assumption we re-calculate our main result by leaving these observations out. The results hardly change, see Table 9. We also use predicted values from this regression (a weighted average of IADL and the memory dummy) as a *disability index*, which we call d.

II.3.b. Imputing missing information for total hours

As already mentioned, our main goal is to estimate how much of total care needs in a given country are covered by different forms of care (NH, FHC, IIC, OIC). A fundamental challenge we face is that neither the HRS nor SHARE record care hours for NH residents. Since our goal is to find which fraction of *care needs* is covered by nursing homes rather than finding the true man-hours employed in nursing homes to take care of a given respondent, the imputations in this section aim to answer the following question: How many hours of care would an individual require if being cared for at home?[15] In other words, we want to make the numbers for NH and community care comparable.

In the following, we describe how we use care hours received by HRS community residents and the observed selection patterns into NH to impute how many care hours a NH resident would have received in the community. An important issue to take into account is that we expect NH residents to be negatively selected in terms of their dependency. Indeed, we find that *conditional on observables* (such as IADL limitations and dementia), the fraction of NH residents is increasing

[15] Most likely, nursing homes need fewer man-hours to meet care needs than community care does due to economies of scale. Estimating this comparative advantage may be of interest, but not the objective of the exercise at hand.

in dependency. It is thus to be expected that negative selection also occurs *on unobservables*. For example, within a group with a fixed profile of limitations (say, a memory limitation and problems with managing money), it is most likely the respondents with most serious limitations (e.g. the most severe dementia) who end up in nursing homes. Thus, even when controlling for observed limitations, we would impute too low a care need to NH residents if we imputed the care need of the NH resident from a community resident with the same observed disability profile. This is a classical selection problem: We only observe the dependent variables (hours) conditional on a discrete choice (community versus NH residence), but unobservables (e.g. the severity of dementia) affect both the discrete choice and the dependent variable.[16]

We now set up our selection model that addresses this issue. Let h_{it} be the *care needs* of individual i at time t, defined as the total daily hours of care that individual i receives (or would receive) at t if taken care of in the community.[17] h_{it} is observed only if the individual chooses to live in the community but not when living in a NH. Let u_{it} be i's utility from entering a nursing home at t; we normalize the utility of staying in the community to zero. To account for observed dependency, we use an *IADL count index* (from 0 to 12) that counts the number of IADL limitations plus the presence of a memory-related limitation; let x_{it} be a vector of twelve dummies that indicate to which of the twelve dependency categories defined by the index a respondent pertains. We specify the following parsimonious selection model that determines h_{it} in the population:[18]

$$u_{it} = x'_{it}\alpha + z'_{it}\gamma + \eta_{it}, \tag{1}$$

$$h_{it} = x'_{it}\beta + \epsilon_{it}, \qquad \text{(population model; } h_{it} \text{ observed if } u_{it} \leq 0\text{)}. \tag{2}$$

Here, z_{it} is a vector of country dummies for Europe and five regional dummies for the U.S. (Northeast, Midwest, South, West, and Others); these constitute our instruments. The coefficient vectors α, β, and γ are to be estimated. ϵ_{it} and η_{it} are mean-zero error terms which we allow to be correlated.[19] We now formally state the assumption we need to impute missing care hours:

[16] We thank an anonymous referee and Jesús Fernández-Huertas for helpful comments regarding this selection problem.

[17] If an individual receives multiple forms of community care (FHC, OIC, IIC), h_{it} is the sum of hours over all forms of care.

[18] We also experimented with richer models, e.g. including the full set of IADL and dementia dummies in Eq. (2). However, this model suffers from a curse of dimensionality, in the sense that the number of cells (i.e. respondents with exactly the same disability profile) quickly multiplies: 2^{12} in this example. It is well-known that the non-parametric selection model we use (Dahl, 2002) gives robust predictions only if there are sufficiently many observed data points for hours in each cell. Also, we found that the fit of the second stage hours regression did not improve much in the richer models (as measured by R^2). We thus opt for the more parsimonious specification given here (which has merely 12 cells).

[19] The correlation of ϵ_{it} and η_{it} is the essence of the selection problem: We expect a positive correlation due to the presence of unobserved components of dependency. Severe dementia, for example, would manifest itself in a positive η_{it}, making NH entry more likely, but also a positive ϵ_{it}, i.e. a higher care hours if taken care of in the community.

Assumption 3 *Given a fixed profile of IADL limitations and dementia, population care hours h_{it} have the same mean in all countries (and all regions of the U.S.), i.e. the following* exclusion restriction *holds:*

$$\mathbb{E}\big[\epsilon_{it}|x_{it}, z_{it}\big] = 0. \tag{3}$$

We assess whether this assumption is empirically plausible through various exercises. In one check, we compare actual (non-imputed) hours of care in SHARE with actual hours of care in the HRS (see Table 15 in the Appendix); we find that hours of care are similar for lower levels of dependency but diverge for higher dependency counts. This pattern is consistent with what we would expect given that selection into nursing homes only becomes operative at higher dependency levels and hence corroborates our assumption. In a second exercise, we make use of the U.S. data and check how well our selection model predicts observed care hours in the community in the U.S. We estimate the selection model (discussed below: "Selection of the benchmark model"). We omit one U.S. region (Northeast, Midwest, etc.) at a time and compare how close predicted care hours are to actual out-of-sample care hours from the omitted region (see Table 14 in the Appendix); we find that the selection model performs well in this task. Finally, we also compare the hours imputations from the selection model to outside-household care hours in Europe (recall that inside-household hours are not observed in SHARE). We find that predicted and actual hours are in the same ballpark, and that predicted hours almost always exceed observed outside-household hours (see Table 16 in the Appendix); this is to be expected since outside-household care recipients can be expected to be positively selected on dependency.

Remark. Despite Assumption 3, we expect *observed care hours among community residents* to differ across countries. In countries with generous subsidies for nursing homes (e.g. Sweden) we expect a larger set of people within each (observed) disability group to be selected into NHs, and within each dependency group those who opt for NH to suffer from stronger unobserved frailty, resulting in a lower observed mean of hours in the community. In countries in which almost all individuals choose to stay in the community (e.g. Italy), we expect higher observed hours among community residents since even those with higher degrees of unobserved frailty often remain in the community.

Instruments. Our identification assumption is that the country or the region (within the U.S.) that the respondent lives in affects only the choice if to go to a nursing home or not, but *not* how many hours of care she receives (or would receive) in the community. We choose this instrument since it is a NH supply shifter: In different regions of the U.S., NH supply is arguably different because the density of NHs vary in the degree of urbanization; nursing homes are sparser in areas with low population density, thus lowering the supply of NH. However, it is reasonable to assume that the lower supply of NH is not correlated with how IC hours in the community are chosen at

the extensive margin.[20]

Estimation. A large class of selection models relies on the insight that one can include a correction function $\mu(\cdot)$ in the second-stage regression that accounts for sample selection (see the survey by Bourguignon et al., 2007). In our case, the second-stage estimation equation for hours is

$$h_{it} = x'_{it}\beta + \mu(P_{it}) + \hat{\epsilon}_{it} \qquad \text{(estimated on U.S. community residents)}, \tag{4}$$

where P_{it} is the probability that individual i is in a nursing home at t (given their IADL count and the region they live in, which is estimated in the first stage) and $\mu(\cdot)$ is a correction function that satisfies $\mu(0) = 0$.[21] We estimate regression equation (4) on the sample of community residents in the U.S. We exclude European countries in the second stage since we only observe hours for the subset of community residents who obtain care from someone *outside the household* and are thus, presumably, twice positively selected; note that it is plausible to expect that the most dependent respondents need a co-resident caregiver.

Our estimation method follows Dahl (2002), who shows that under Assumption 3, OLS yields consistent estimates of the parameter vector β when modeling $\mu(\cdot)$ non-parametrically, using for example a polynomial. In Eq. (4), we obtain the estimate for P_{it} from a first-stage probit estimation of Eq. (1), which we run on the sample of all care recipients in both SHARE and the HRS. In the second stage, we opt for the linear specification $\mu(P_{it}) = \delta P_{it}$ since in practice, the fit of the second stage does not improve when going beyond a linear term. This parsimonious model also has the advantage of being easily interpretable.

Our estimation results are consistent with NH residents being negatively selected. The coefficient estimate on P_{it} is $\hat{\delta} = -5.46$, meaning that a 10-percentage-point increase in the NH-probability (for a group with fixed dependency profile) lowers care hours observed among community residents by half an hour per day. This estimate implies sizable adjustments in the imputation

[20] For a further discussion of the instrument, see Paragraph "Robustness" below. We also considered other instruments, e.g. variables that correlate with the supply of *informal care*, such as the respondent's coupledness status, the number of children, and a daughter dummy. However, these variables are likely to violate the exclusion restriction: We expect them to have an influence both on the extensive and intensive margin of informal care. For example, the presence of spouse in the household should make both NH entry less likely and lead to higher care hours since the spouse's cost of caregiving is low due to his/her spatial proximity. In fact, regressions indicate that the availability of caregivers (spouses, children, daughters) increases hours of care received.

[21] The workings of the correction function $\mu(P_{it})$ can be understood from the following stylized example. Suppose there was a region in which all respondents are observed in the community for a given IADL count, say $x_{it} = 2$. Then, we would have $P_{it} \simeq 0$ in the first-stage for this group in this particular region. The second-stage estimator would then use this region to identify the population mean of hours for respondents with two limitations, β_2, since we constructed the correction function such that $\mu(0) = 0$. If in a second region the NH-probability is 0.5 for the same dependency group, the estimator will use this information to infer how the selection of half the population into NH decreases the observed hours in the community below the true mean. If there is no region with a zero NH probability, the estimator achieves identification by extrapolating the effects of selection towards $P_{it} = 0$, using the exogenous variation of the NH probability across regions.

of care hours for NH residents. With respect to a naive OLS imputation scheme, our estimator assigns hours that are between five and eight hours higher, depending on the country in question. The corrections are smallest in the Northern countries, where the fraction of care recipients selecting into NHs is large, but the correction increases as we move South, where only a few elderly (presumably the most severely dependent) are in NHs, see Table 11 in the appendix. Thus, correcting for selection somewhat attenuates the North-South gradient in NH-provided care.

For the SHARE countries, we impute IIC hours in a final step. We use our estimate for average care hours provided to community residents and the observed hours for outside-household care (OIC) to infer IIC hours. We refer the reader to the appendix for details.

Robustness: Discussion and alternative imputation schemes. We note at this point that the identification assumption for our instrument (U.S. regions) may be debatable. For example, rural populations may have stronger family ties and thus also choose higher IC hours at the extensive margin. In order to check the robustness of our main results, we thus consider two alternative imputation schemes, which we describe in more detail in the appendix:

(1) **Pooled OLS**: We run an OLS regression of care hours on the set of IADL dummies, the memory dummy, and age. We then use predicted values from this regression to impute hours whenever they are not observed, i.e. for NH residents in all countries and for IIC recipients in SHARE. We note that this estimator neither addresses the issue of non-random selection of NH residents nor does it exploit the information whether an inside-household caregiver is present.

(2) **Grouped OLS**: We run the same regression as in pooled OLS, but do this separately for the IIC and OC samples in the U.S. We use predicted values from the IIC regression (OC regression) to impute hours for IIC recipients (OC, in case that they are missing) in SHARE. For NH residents, we do not know if they would have received OC or IIC had they opted to stay in the community; we thus adopt an agnostic stance and impute care hours from a regression on the joint sample (which equals the pooled-OLS imputations). This specification takes into account the presence of an inside-household caregiver (which likely correlates with more severe dependency); however, we note that it does not address non-random selection NH residents, either.

Selection of the benchmark model. There is no way to test how well the different imputation schemes predict hours for NH residents, since their care hours are never observed. However, we can get a sense of how well the different schemes perform to impute community hours in the HRS, which we do observe. In a horse race between the specifications, we exclude one of the five regions in the U.S. at a time and act as if we did not observe hours in this region. We then estimate the three models on the remaining four U.S. regions and impute community hours for the omitted region.

We then compare how close the imputations of the three schemes come to the average of observed hours.[22] Table 14 in the appendix shows that the Dahl selection model performs best. This horse race also lends credit to our assumption that care hours follow the same distribution conditional on IADLs (as we have briefly mentioned above after introducing Assumption 3). Nonetheless, the other two models do not perform much worse; grouped OLS actually performs best when restricting attention to the four mainland regions of the U.S. This finding suggests that IADLs do a good job in what they are designed to do, which is measuring care needs.

II.4. Splitting total care hours into different categories

For community residents, we have so far only discussed how we impute *total* care needs in hours, but not the actual forms of care (IIC, OIC, FHC) a person obtains.[23] [24] For IIC recipients in SHARE, this requires an additional step. For more than half of this group, individuals receive only informal care and thus the assignment is unproblematic. In the remaining cases, we subtract care hours in known categories (FHC and/or OIC) from the total imputed hours to obtain IIC hours. We perform similar imputations for observations with missing hours that are due to other reasons, but these are not quantitatively important for our main result; see the appendix for more details.

II.5. Other issues

Finally, there are other more minor challenges in SHARE that are worthwhile to point out. Firstly, for co-residing couples answers to the question on who received care from whom are often contradictory. In order to resolve this inconsistency, we use data on (I)ADL limitations to determine the recipient; this is straightforward in the large majority of cases. Secondly, using HRS data we impute hours for meals-on-wheels (food delivered to the elderly's home by organizations) which is absent in SHARE.

III. Institutional features of public LTC

In order to put our empirical results across countries into perspective, we first provide a brief descriptive overview of the central features present in publicly provided LTC benefits. Table 4

[22] We choose this criterion since country means are what we ultimately care about in our main results.

[23] Depending on the policies in different countries, these choices may well differ. For example, Northerners may cover a larger fraction of their care need by FHC if it is more readily available or because of higher subsidies than in the South.

[24] For NH residents, we set care hours of all other care forms to zero. NH residents may receive visits from relatives who perform certain tasks, but this is likely negligible compared to care received from nurses.

summarizes the most pertinent characteristics for our countries. The ordering of countries is descending in accordance with their share of GDP devoted to public LTC spending in 2015.

An immediate observation is that public spending on LTC coverage varies widely across our countries. As a benchmark of comparison we note that the average public LTC expenditure as a share of GDP across OECD countries is 1.7% (not shown here – see Fig. 11.24 in OECD, 2017). The ordering of the countries reveals that across the European countries, expenditure shares follow a North-South gradient with high public spending in Northern Europe and low spending in Southern Europe. The Netherlands and Sweden, for example, spend almost double the amount of their GDP on LTC than the OECD average, whereas, public spending in Spain and Italy is less than half of the OECD average. The lowest public spending on LTC as a share of GDP is in the U.S. with only one-third of the OECD average.

Furthermore, countries display significant variation in how LTC benefits are provided. LTC benefits can predominantly be in the form of formal in-kind services at home or in a nursing institution (e.g. Sweden, Denmark), provided as a cash benefit (e.g. Austria, France), or the person in need of care is given a choice between in-kind services and cash benefits (e.g. the Netherlands, Germany). In cases where cash benefits are offered directly to the care recipient, they can generally be used for formal care at home or in a nursing home and to compensate informal caregivers (e.g. Germany and Spain). However, cash benefits tend to be substantially lower than the value of the corresponding in-kind service. For instance, in Germany and the Netherlands cash benefits are, respectively, 50% and 25% lower than when opting directly for the assistance. Cash benefits also come with more-or-less stringent restrictions, such as in Spain, where informal caregivers can be compensated only in regions where the responsible administration is unable to provide formal care services. In Denmark, care recipients cannot use the cash benefit to compensate informal caregivers for providing nursing care (e.g., help with ADLs) but only for help with activities such as gardening or financial tasks (e.g., help with IADLs) and only if it serves as a complement to formal home care, provided that formal care takes place for at least 20 hours per week. In France, cash benefits can be used to compensate informal caregiving but not if it is done by the spouse or the partner; they are legally required to provide care. In Germany, recipients have considerable leeway in the compensation of informal caregivers.

In most countries, access to public LTC benefits is *universal* but benefits can differ strongly depending on the individual's circumstances. Specifically, in countries with universal LTC schemes, anyone who has sufficient care needs, and in some cases belongs to a certain age group, is eligible for benefits. Here, the U.S. is a stark outlier where access is *means-tested* and the test is more akin to qualifying for welfare benefits. That is, a dependent elderly is only granted access to government-provided support for LTC if income is sufficiently low and assets are essentially exhausted. However, even where access is universal, benefits are typically insufficient to cover the

TABLE 4: Overview of public LTC programs

Country	Program(s)	(1) Gov't LTC spending/GDP	(2) In-kind/cash/ both	(3) Access, ages	(4)% of LTC exp. out of pocket	(5) Cash benefits (monthly in EUR)
Netherlands	AWBZ	3.7	both	universal, all	0.0%	1,250-1,500
Sweden	Ädelreformen	3.2	in-kind	universal, all	0.8%	100-300
Denmark	CASS	2.5	in-kind	universal, all	10.4%	as complement to FHC
Belgium	INAMI	2.3	in-kind	universal, all	10.0%	
	Medc.-leave Allow.		cash	universal, all		741 (flat)
	Assistance Allow.		cash	means-test, 65+		77-516
	Flemish Insr.		cash	universal, all		130 (flat)
France	APA	1.7	cash	universal, 60+	0.8%	530-1,235
Germany	Soz. Pflgvrsg.	1.3	both	universal, all	30.4%	225-1,750
Austria	BPGG	1.2	cash	universal, all	17.1%	154-1,655
Spain	LAPAD	0.8	both	means-test, all	28.1%	400-831
Italy	RSA, IDA	0.7	in-kind, cash	universal, all	unknown	472 (flat)
U.S.	Medicaid	0.5	in-kind	means-test, 65+	40.0%	some states

Netherlands: AWBZ (Algemene Wet Bijzondere Ziektekosten): Cash benefits are provided by the program 'Persoonsgebonden budget' and are 75% of the value of the corresponding in-kind service. 12% of carees opt for the cash option. **Sweden**: Ädelreformen: Cash benefit is meant to be symbolic only. **Denmark**: CASS (Consolidation Act on Social Services): includes cash benefits for help with activities that complement nursing care at home, e.g. help with gardening, financial tasks; based on at least 20 hours of weekly care. **Belgium**: INAMI (Institut national d'assurance maladie-invalidite) is the national health insurance system and is primarily responsible for LTC. There are several supplementary programs, such as the medical assistance leave allowance of up to 12 months directly targeted at the family caregiver, home-help assistance (in-kind), and an allowance for assistance of the elderly which is income- and asset-tested for ages 65+. On the regional level, there is the Flemish care insurance restricted to Flemish residents. **Austria**: BPGG (Bundespflegegeldgesetz), cash benefits to buy formal care services or to reimburse informal caregiving. **France**: APA (Allocation personnalisee d'autonomie) while not means-tested, decreases very rapidly in income (from 0% to 80%). It can be paid to informal caregivers but not to the spouse or partner. **Germany**: Soziale Pflegeversicherung (universal LTC insurance): Cash benefits are about 50% lower than the corresponding in-kind services at home and can be used to compensate informal caregivers. 52% of carees opt for cash. **Spain**: LAPAD (Ley de Promocion de la Autonomia Personal y Atencion a las Personas en Situacion de Dependencia). The dependent person is entitled to financial assistance only if the competent administrations are unable to offer these services. There are three types of financial benefits, all income-tested: to purchase formal care services, to compensate informal caregivers, to hire a formal personal caregiver. There have been considerable delays between claiming benefits and actually receiving them. **Italy**: RSA (Residenze Sanitarie Assistenziali): Institutional care is part of health system. IDA (Indennita di accompagnamento) is an attendance allowance and comes with considerable flexibility. **U.S.**: Medicaid is the primary payer for LTC services designed to assist people with limited income and assets. Means-testing is very strict.

Sources: (1) OECD (Health at a glance 2017 – OECD indicators); numbers refer to year 2015 or nearest available. (4) LTC expenditures paid out of pocket (note, private insurance in Belgium accounts for 9.8% and is included here); OECD (2011, Help Wanted? Figure 7.3 and Figure 1.8); numbers refer to year 2007/2008 based on OECD Health System Accounts, 2010. The number for Italy is not available. The number for the U.S. is based on Figure 1.8 of the OECD article. OOP does not include board and lodging in a nursing institution. (2), (3), and (5) are based on OECD (2011, chapters 1, 4 and 7) and country-specific reports from the European Network of Economic Policy Research Institutes (2010).

entire cost of care. Thus, despite the fact that LTC coverage in all countries is predominantly publicly funded, private (out-of-pocket) spending can be high. Northern European countries tend to have the lowest out-of-pocket expenditure followed by Central and Southern European countries, the U.S. exhibiting the largest share of private expenditure. The generosity of the benefits – and so the extent of out-of-pocket expenditures – often depends on income (less often on assets) of the care recipient, in addition, to the severity of the dependency. In France, for example, cash benefits decrease steeply in income. In Spain, all programs are income-tested. In some cases, even income and assets of the care recipients' families matter in the determination of benefits (e.g. Germany and France). In contrast, some programs do not tie benefits to income or assets and instead provide a flat allowance to the care recipients. This is, for example, the case with the attendance allowance in Italy ('Indennita di accompagnamento'). Public expenditure shares therefore potentially mask the extent of insurance that public LTC programs provide. For example, while Italy and Spain have comparable shares of public LTC expenditure of GDP, the policy in Spain is targeted to those who need the financial support most, which suggests that Spain's expenditure on LTC has a higher insurance value than the equivalent expenditure has in Italy. Finally, in all countries, at least in the absence of complete impoverishment, public funds do not cover the entire cost of room and board in a nursing home. Northern European countries (especially the Netherlands and Sweden) are again more generous in this respect, whereas others do not cover any of these expenses such as, the German LTC program.

Finally, while countries typically provide benefits directly to care recipients, benefits for informal caregivers have also become increasingly commonplace, even in countries that have traditionally considered elderly care to be the responsibility of the state, like Sweden and the Netherlands. Often these benefits are not allowances *per se*, but rather allow for carers to maintain employment and mental health. Policies include, but are not limited to: leave from work, flexible work hours, care leave, respite care, counselling and training services, and information and coordination services for carers. Leave from work can be paid or unpaid. Belgium, for example, provides paid medical leave for up to 12 months, longer than any other country; Germany provides unpaid leave up to six month; Spain provides unpaid leave of up to three years (in extreme cases). Benefits for informal carers are still less prominent in Northern European countries than in Central and Southern European countries and are non-existent in the U.S. on a national level (some states have experimented with cash benefits for care, e.g. California and Florida).

We conclude this section by noting that even though countries differ markedly in their approach to LTC (and some even differ within, e.g. Belgium), countries can be broadly clustered by geographic regions based on central commonalities along several dimensions. Doing so will at times prove useful for our empirical analysis in order to increase the size of the sample when zooming into more specific details. We group European countries into three geographic regions

and treat the U.S. as a separate group. *North* comprises countries whose public spending on LTC exceeds the OECD average, provide more generous benefits, which are more often in-kind than in cash, and who have traditionally placed the greatest emphasis on the state in the responsibility of elderly care; these countries are in Scandinavia and the Benelux. *Middle* countries are defined as those with public LTC spending near and below the OECD average but above 1% of GDP, where cash benefits are the more prevalent form of public support; Middle countries are overall less generous than countries in the North and have placed a greater emphasis on the importance of informal caregiving; all are found in the center of Europe. The remaining countries comprise the region *South*. Their public LTC spending is comparatively small and their formal LTC sector is much less comprehensive.[25] Finally, we define the U.S. as one region since its LTC policy, specifically the stringent means-testing of assets and income to determine eligibility, make it very different from LTC policies in European countries.

IV. Empirical results

The main purpose of this section is to provide a quantification of the relative importance of all forms of elderly care by country. In order to do so, we use our data to calculate relative frequencies of care arrangements and typical hours of care a person aged 65 and older receives. We always use as many waves as possible as long as this allows us to obtain consistent estimates. For statistics on care hours in SHARE, for example, this means that we typically have to restrict the sample to Waves 1 and 2 since only these waves have sufficient information to infer the distribution of hours.[26] For several other questions, however, we are also able to use later waves.[27]

IV.1. Demographics and dependency rates

Before turning to quantifying care arrangements, it is useful to gain an understanding of the demographic situation and the rate of dependency across countries. Table 5 shows a summary of these features ordered by the share of public spending of GDP on LTC programs.

The first data column shows the ratio of the retired population (above 65 years old) over the working-age population (20 to 65 years old). We see that there are no systematic differences

[25] We note that our categorization of countries is also in line with the classification of the European Commission (2015) which is based on qualitative features of countries' policies.

[26] Our imputations give us hours information for many respondents also in Waves 4 through 6, but the sample of such respondents is biased towards respondents who receive only one category of care.

[27] Wave 3 has no information on current care arrangements for respondents residing at home; we only use Wave 3 when it comes to NH status but exclude it for all other exercises. Wave 4 contains no information on formal home care; we thus exclude Wave 4 when calculating any statistics involving formal home care but use it when it comes to statistics on informal care.

between countries in Europe, but that the U.S. has a substantially lower dependency rate. The second column shows that the fraction of over-65-year-olds who suffer from severe dependency, defined as those in need of 3 hours of daily care or more (i.e. $d \geq 3$), shows little variation across countries including the U.S. The reason for why the U.S. has a similar dependency ratio as the European countries but is younger stems from an onset of heavy LTC needs at younger ages. Figure 1 displays our dependency ratio by age groups. Among 65 to 69-year-olds the dependency ratio in the U.S. is largest and about twice as high as for the Middle region. Interestingly, European countries catch up and surpass the U.S. in the oldest age category. Thus, the similar dependency rates for *all* above-65-year-olds mask substantial heterogeneity when it comes down to age groups.

TABLE 5: Demographics, dependency, and government spending in 2015

Region	Country	Dependency ratio in % (65+/20-65)	Disability ratio (% of 65+ with $d \geq 3$)	Gov't LTC spending/GDP
North:	Netherlands	30.2	9.6	3.7
	Sweden	33.8	9.8	3.2
	Denmark	33.0	9.6	2.5
	Belgium	30.6	14.0	2.3
Middle:	France	33.3	11.0	1.7
	Germany	34.8	11.0	1.3
	Austria	30.5	9.9	1.2
South:	Spain	30.6	14.7	0.8
	Italy	37.8	11.5	0.7
U.S.:	U.S.	24.6	11.9	0.5

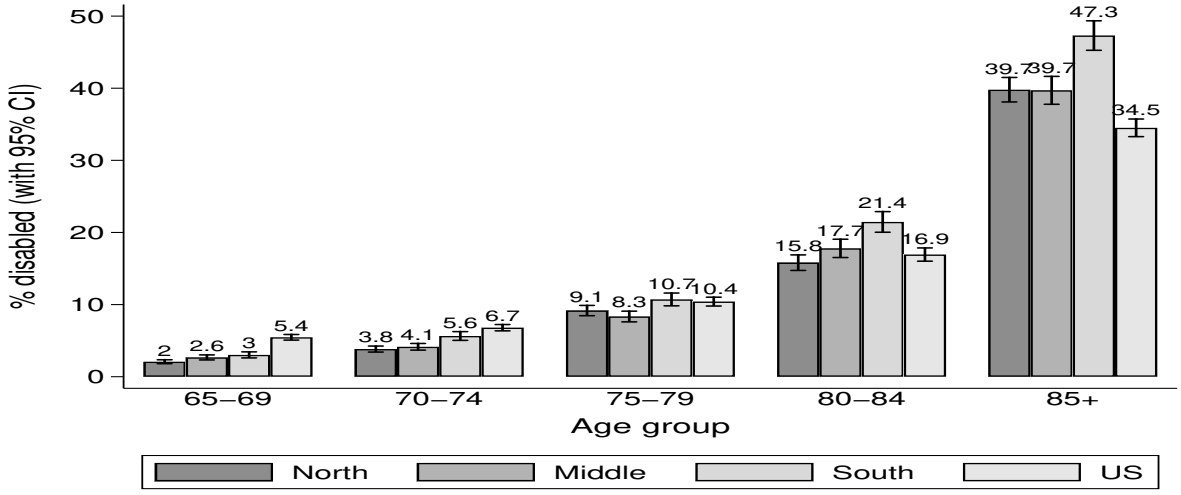
Dependency ratio is taken from United Nations (2017), LTC spending ratio from OECD (2017). Dependency is for the year 2015, LTC spending ratio is for the year 2015 (or nearest year available) and includes both health LTC spending (mainly ADL-related) and social LTC spending (mainly IADL-related). The disability ratio is obtained pooling all waves in SHARE but Wave 3 (no disability data); for the U.S. we pooled all waves after 2003 from the HRS. The disability ratio is calculated as the fraction of individuals aged 65 years and above who have a disability score d (daily care need in hours) of at least 3.

Another key message Table 5 provides is that there is no apparent association between public spending and disability rates. If anything, governments in the Northern and Middle countries, which have somewhat lower care needs, spend most on LTC. This begs the question whether countries with less public insurance against LTC risk display higher rates of private LTC insurance which makes up for the deficit. Figure 2 shows that in general very few people take up private LTC insurance unless they are forced to do so (which is the case in some countries). Only Belgium, France, and the Netherlands show substantial sign-up rates, but even in these countries fewer than one-fourth of over-65-year olds has voluntary LTC insurance.[28] Thus, there is no evidence that

[28]Only Waves 5 and 6 in SHARE asked explicitly about LTC insurance; the Netherlands did not participate in Wave 6. From Wave 5 to Wave 6, SHARE changed the feasible set of answers to question `hc116` in many countries, probably because the framing of the question had led to misunderstandings in its original form. We opt for the latest wave since it should have the best data quality.

people use private insurance markets more in countries in which the government spends little on LTC (such as the South and the U.S.). Instead, as we will see below, the shortfall in government spending is compensated by informal caregiving.

FIGURE 1: Cross-region disability ratio by age group



The disability ratio is calculated as the fraction of individuals per age group within a region who have a disability score d (daily care need in hours) greater or equal to 3. For North, Middle, and South we use all waves in SHARE but Wave 3 (no disability data); for the U.S. we use all waves from the HRS after 2003. The error bars are 95-percent confidence intervals.

IV.2. The relative importance of care arrangements

We now turn to our main result: the quantification of care arrangements across countries and regions.

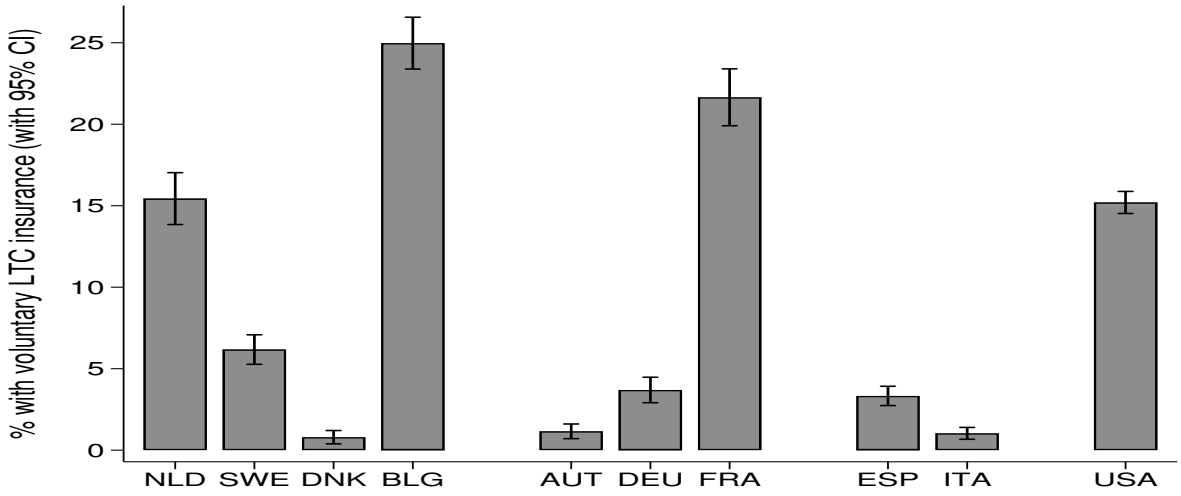
We first present simple case counts. Table 6 shows the percentage of the frail elderly which are taken care of at home informally by family members or friends (IC), which percentage receives formal home care (FHC), and which percentage resides in a nursing home (NH). Additionally we report the percentage of individuals who receive care at home from both informal and formal caregivers in the column *Mix IC-FHC*.[29] The main finding of Table 6 is as follows,

Result 1 *There is a strong North-South gradient in the utilization of informal care: Southerners use IC most and Northerners least, the U.S. being most similar to the South.*

The gradient in nursing-home (NH) and FHC utilization is exactly the opposite, although the U.S. is closer to the Middle than to the South in NH utilization.

[29]We note that Table 6 reports case counts and so is unaffected by any imputations of care hours. For the percentage of the NH population we use the weights constructed by us in line with the OECD reported nursing-home statistics to correct for the under-sampling of this population in SHARE.

FIGURE 2: Voluntary LTC insurance uptake as of 2015



Sample: all respondents aged 65 and above. SHARE: Wave 6 (Netherlands: Wave 5); HRS: Wave 11 (2012). Weights are used.

Another noteworthy observation from Table 6 is that variation in care arrangements within regions is small in comparison to variation in care arrangements across regions. This hints at the possibility that public policy plays an important role in the determination of care arrangements. In this regard, differences in care arrangements between Northern and Middle countries are especially meaningful as these are culturally more alike. The Netherlands and Germany, for example, are culturally similar, yet the frequency of IC in Germany is more than double that of the Netherlands. In line with this, the Netherlands have double the amount of NH residency. Indeed, Bakx et al. (2015), using a decomposition exercise for the probability of care choices, argue that differences between these two countries are primarily due to differences in their institutional setting. On the other hand, differences in care arrangements between France and Belgium are less pronounced, weakening the case for the importance of policy in favor of culture (social norms, religiosity) in determining care outcomes. Nonetheless, nursing home utilization in Belgium is well above that in France and even exceeds Denmark's.

Table 6 shows that there is a substantial number of elderly who live at home and obtain both informal and formal care in all countries. This seems to suggest that combining the two care forms is an attractive option. However, we will now see that once we account for the intensity of care, one of the two care forms tends to dominate at the individual level.

In order to quantify the relative importance of IC and FHC for those who use both, we classify respondents into one of three categories based on the contribution of IC to total care hours. In the first category, IC accounts for the vast majority of care hours (at least 80%), the second consists of "true mixers" (20-80% of IC), and in the third IC plays a minor role (less than 20%). Figure 3

TABLE 6: Care arrangements across countries (in % of cases)

Region	Country	Informal care (IC)	Mix IC-FHC	Formal home care (FHC)	Nursing home (NH)
North:	Netherlands	14	28	17	42
	Sweden	27	27	11	35
	Denmark	21	37	15	27
	Belgium	20	35	14	31
	NORTH	20	32	14	34
Middle:	France	24	42	14	20
	Germany	36	35	7	21
	Austria	36	38	12	15
	MIDDLE	31	38	11	20
South:	Spain	57	26	8	8
	Italy	62	22	8	7
	SOUTH	60	24	8	8
U.S.:	U.S.	64	12	5	19
All:	ALL	54	21	7	18

Sample: All respondents aged 65 and above who receive care. SHARE: Waves 1, 2, 5, and 6. HRS: Waves 5-11. Weights are used.

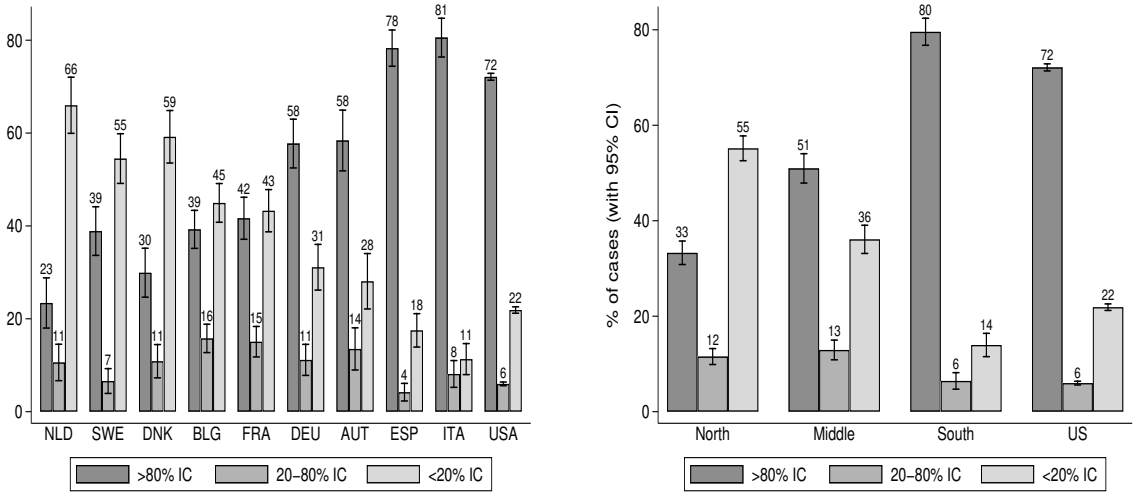
displays the results. In all countries, true mixers are in the minority, accounting for less than one-fifth of all who combine the two care sources. The tilting is strongly towards IC in the South and the U.S., more balanced in the Middle, and tending towards formal care in the North, in line with the general tendency that we identified in Table 6. This finding also lends support to the literature which finds that informal and formal care are rather substitutes than complements.[30] We summarize this result as follows:

Result 2 *While IC and FHC are often complementary, their relative importance is attributable to one form of care: IC in the South and the US, FHC in the North, and IC or FHC in the Middle.*

Table 6 further seems to suggest that in all countries care given at home (IC and FHC) is more important than NH care. However, the case counts certainly overstate the importance of care at home since it is individuals with the highest levels of dependency who tend to opt for nursing homes. We now correct for care intensity by measuring the average daily hours of care that a typical individual of age 65 years and above receives in the four regions (note that averages are taken over the *total population* above 65, i.e. including those without care needs).

[30]Note that this result does not hinge on the imputation of hours for IIC. As a robustness check, we computed the same statistics as in Figure 3 for the sample to those who did not receive IIC. The percentage of those receiving between 20 and 80% IC stayed within a range of 3 percentage points compared to the results in Figure 3.

FIGURE 3: Choices of IC-FHC mixers



Sample: All respondents aged 65 and above who received both IC and FHC. SHARE: Waves 1-2. HRS: Waves 5-11. Weights are used.

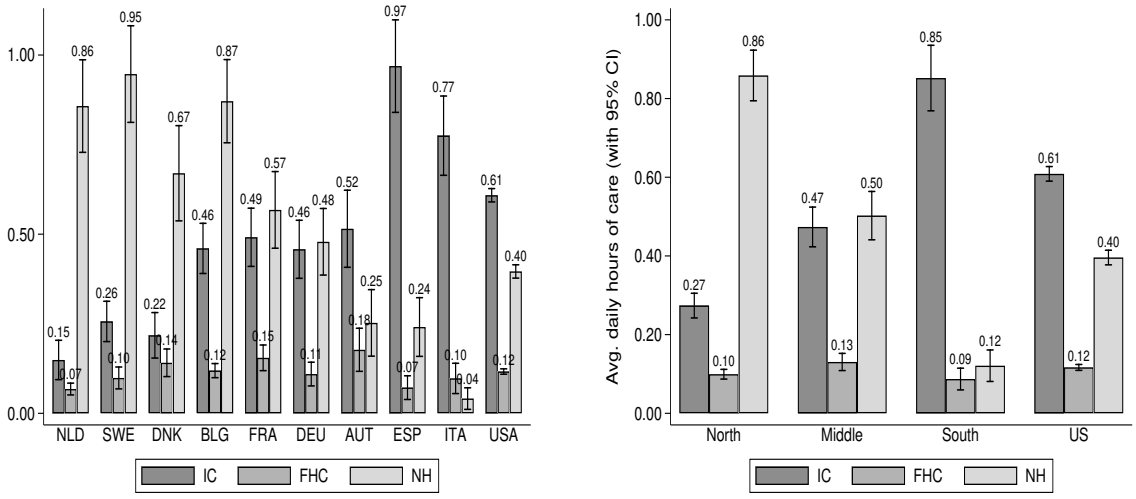
Figure 4 shows the results. We see that the importance of NH dramatically increases relative to the case counts in Table 6, especially in the North, where NH now exceeds IC and FHC. Overall, FHC becomes less important but IC remains prominent, indicating that IC is also used in cases of more severe dependency. Finally, the North-South gradient document in Table 6 remains robust even when accounting for the intensity of care. The U.S. is now positioned between the South and the Middle, but still far away from the North.[31]

Result 3 *When accounting for the time intensity of care, IC and NH emerge as the most important care forms. The North-South gradient in IC stays very strong, the U.S. falling between the European Middle and South.*

When comparing the variation in care arrangements with the variation in LTC policies, a stark association becomes discernible: More generous public LTC insurance comes with lower IC and higher NH. A possible chain of causality could be from policy to care arrangements. More generous public LTC benefits induce people to favor formal sources of care. Under this reading, our results suggest that policy is very effective in influencing care decisions. An alternative interpretation is that culture is the main cause behind care arrangements and that policies merely reflect societies' values. This reading would be consistent with policy being less effective in shaping behavior, as people operate according to rules (norms) which are difficult to alter over shorter periods of time. Assessing the relative importance of policy and norms or other cultural factors

[31] This result is robust to the choice of the imputation scheme, see the discussion in Section V..

FIGURE 4: Hours of care in different care forms



Sample: All respondents aged 65 and above. SHARE: Waves 1-2. HRS: Waves 5-11. Weights are used.

in driving these cross-country patterns is outside the scope of this paper and requires a different methodological approach.[32]

Undoubtedly, IC constitutes an essential source of care, especially where public-sector provision of LTC is limited. However, as with any non-market traded service, the value of IC does not enter GDP. We now use our care-hours estimates to impute the market value of IC. In order to do so, we need to find a wage rate which reflects the "price" of IC. An obvious choice is the wage rate of an unskilled home-caregiver; these wages are, however, difficult to obtain for our countries. Instead, we provide a lower bound by making use of minimum wages and collective bargaining agreements in countries for which a legal minimum wage is absent. Since caregiving is a low-skill task, these lower bounds are likely close to the true market value.

Table 7 shows that the large variation across countries in public LTC expenditure is substantially reduced when adding the imputed value for IC (last column); this is because IC as a share of GDP has a steep North-South gradient. For Southern countries, the IC share of GDP even exceeds the public expenditure share (the latter being 0.8% for Spain and 0.7% for Italy). In the Central countries, the IC share is about half of the public expenditure share (public shares: France 1.7%; Austria 1.2%; Germany 1.3%). In the U.S., the IC share of GDP is close to that of the public sector

[32]Bakx et al. (2015) compare the Netherlands and Germany and find that institutional differences, especially differences in eligibility rules, are crucial in explaining the higher use of formal care in the Netherlands than in Germany, two countries which are culturally similar and so lends support to the importance of institutions. Dobrescu (2015) uses a structural life-cycle model to study the savings behavior of the elderly in Europe and finds that the elderly have a strong incentive to keep wealth for strategic reasons (to induce family provide care) which interacts with family cohesion which is stronger in Mediterranean countries than in Central European and Scandinavian countries. This finding lends support to the notion that informal caregiving also depends to an important extent on social norms.

TABLE 7: Valuation of IC

Country	Min. wage ($ 2017)	GDP per capita ($ 2017)	Elderly share of total population	IC value per capita ($ 2017)	IC/GDP	IC+Gov't LTC spending/GDP
Netherlands	10.1	48,223	17	94	0.2	3.9
Sweden	10.3	53,442	20	195	0.4	3.6
Denmark	17.6	56,307	18	254	0.5	3.0
Belgium	10.0	43,323	18	302	0.7	3.0
France	11.0	38,476	18	354	0.9	2.6
Germany	10.0	44,469	21	352	0.8	2.1
Austria	7.5	47,290	18	256	0.5	1.7
Spain	5.4	28,156	18	344	1.2	2.0
Italy	8.0	31,953	21	470	1.5	2.2
U.S.	7.3	59,531	14	226	0.4	0.9

Sources: Minimum wage and the fraction of the elderly (age 65 and above) of the total population are taken from OECD.org for the year 2017. Sweden, Denmark, Austria, and Italy have no official minimum wage but collective bargaining agreements are in place which stipulate a minimum wage; the information for these countries is taken from the US Department of State for 2017. GDP per capita numbers taken from Worldbank.org for the year 2017. Min. wage is hourly, GDP and IC values are 2017 $ per year. The IC value per capita is obtained as elderly share $\times$ IC hours received by typical 65-year-old (see Fig. 4) $\times$ min. wage $\times$ 365.

(0.5%). In the North, the contribution is minor compared to what the government does, maybe with the exception of Belgium where IC is about one-third as large (Netherlands 3.7%; Sweden 3.2%; Denmark 2.5%; Belgium 2.3%).

IV.3. Additional results

We now use the assembled data set to study how care arrangements in our countries vary along key dimensions, broadly, over time, as well as by the severity of dependency and the marital status of the care recipient. Severity and marital status turn out to be especially informative about the care arrangement an elderly person is likely to be in. We then direct our focus towards describing the characteristics of the informal caregivers who bear the main burden of caring for the elderly. In what is to follow, we present our results grouping over regions rather than over countries, since sample sizes become quite low when we zoom in on specific subgroups. We relegate the country-level counterparts to the appendix when appropriate. Furthermore, now our sample will always be confined to care recipients.

IV.3.a. Time trends of care arrangements

Some of the countries have undergone important changes in their demographic and family structure in the last decades. While our data is insufficient to provide a complete historical picture, SHARE

and the HRS have enough waves so that we can learn about general tendencies of the provision of care recently. Figure 5 shows the percentage of the sources of care for the four regions over time. We restrict ourselves to case counts since SHARE's Waves 5 and 6 do not have hours information on care. We summarize the key message as follows:

Result 4 *In Europe, IC use has decreased at the expense of NH. This trend was strongest in the Middle and South. In the U.S., however, there was a weak trend in the opposite direction.*

Ageing populations and changes in the family structure may be an important source for these changes over time. The finding that in the U.S. the trend in IC has been (slightly) increasing may stem from the fact that its population has been relatively younger than the population in Europe, and that the quickly growing cost of formal care in the presence of stringent means-tested policy has made family-provided care more attractive, a force that is less pronounced in Europe. The trend towards NH in Europe would have presumably been even stronger in the South had the economic crisis not hit these countries with exceptional power. High unemployment certainly lowered the opportunity cost of IC substantially in the South for many families in the last ten years. Nevertheless, the data show that there was a 10-percentage-point decrease in IC over our sample period. This seems to indicate that other long-term trends (such as the incorporation of more women into the labor force) were apparently stronger than the economic crisis in driving care choices.

IV.3.b. Severity of dependency and the care choice

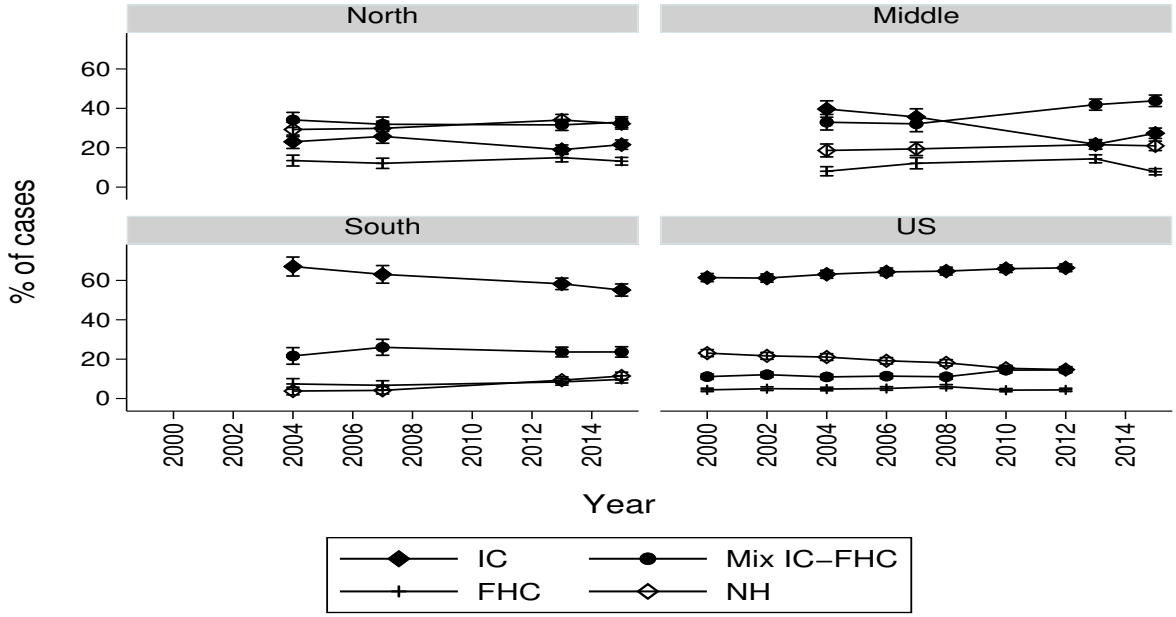
Unsurprisingly, the severity of care needs plays an important role in the care choice in all of our countries. Figure 6 plots the source of care against the disability index, d. The following result emerges:

Result 5 *In all regions, IC use decreases and NH use increases with the severity of dependency. This dependency-IC gradient is strongest in the U.S. and weakest in the Middle and the South.*

On the flip side, it is worthwhile to point out that IC appears feasible even among those who display the highest levels of care needs. Especially in the South and, to a lesser extent, in the U.S. we can see relatively large fractions of heavily-dependent persons receiving the majority of care from informal sources.

Furthermore, we find that the lion's share of care hours is received by those with the heaviest care needs. Figure 7 demonstrates this by plotting the *Lorenz Curve* of care. Each point on the curve corresponds to one point on the cumulative distribution function (cdf) of hours (across all

FIGURE 5: Time trends by region



Sample: All respondents aged 65 and above who received care. SHARE: Waves 1, 2, 5, and 6. HRS: Waves 5-11. Weights are used. North excludes Netherlands (no data in Wave 6).

forms of care). The point marked with $d = 3$, for example, indicates that close to 60% of individuals have care needs of 3 hours and below (the x-component) but account for less than 20% of all hours of care (the y-component).[33] The following finding emerges:

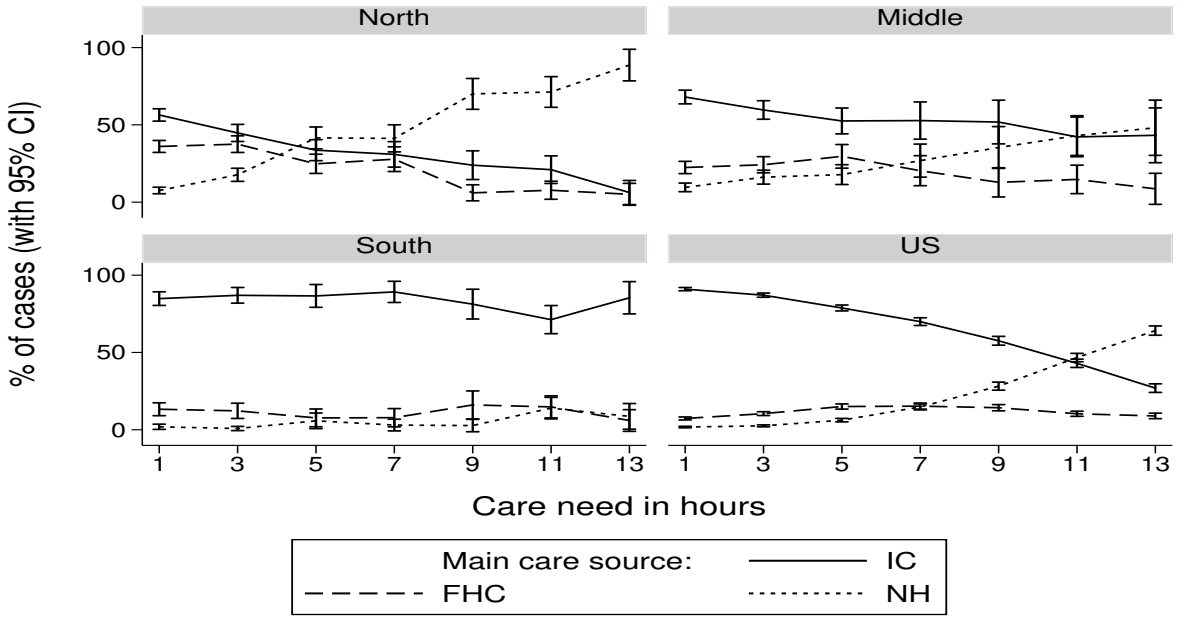
Result 6 *Care hours are concentrated. The lion's share of care hours is due to a relatively small fraction of individuals with heavy care needs.*

IV.3.c. Marital status and the care choice

We now study how care arrangements across countries differ depending on whether the person in need of care is single or coupled, which is naturally related to the availability of informal caregivers. Figure 8 shows the results. For coupled individuals, IC is the predominant form of care in all countries. What is remarkable is that in the North, NH is substantial even among those who are partnered, presumably reflecting the relative desirability of NH care in these countries; this is especially the case in Belgium and Sweden, as Figure 12 in the appendix shows. What stands out among singles is that IC continues to far outweigh NH in the South and remains very significant in the U.S. Also in the Middle IC plays an important role for singles (in Austria IC even dominates

[33] Fig. 7 still –most likely– underestimates the true concentration of care hours since we use predicted hours and not actual hours of care.

FIGURE 6: Care choice by disability level



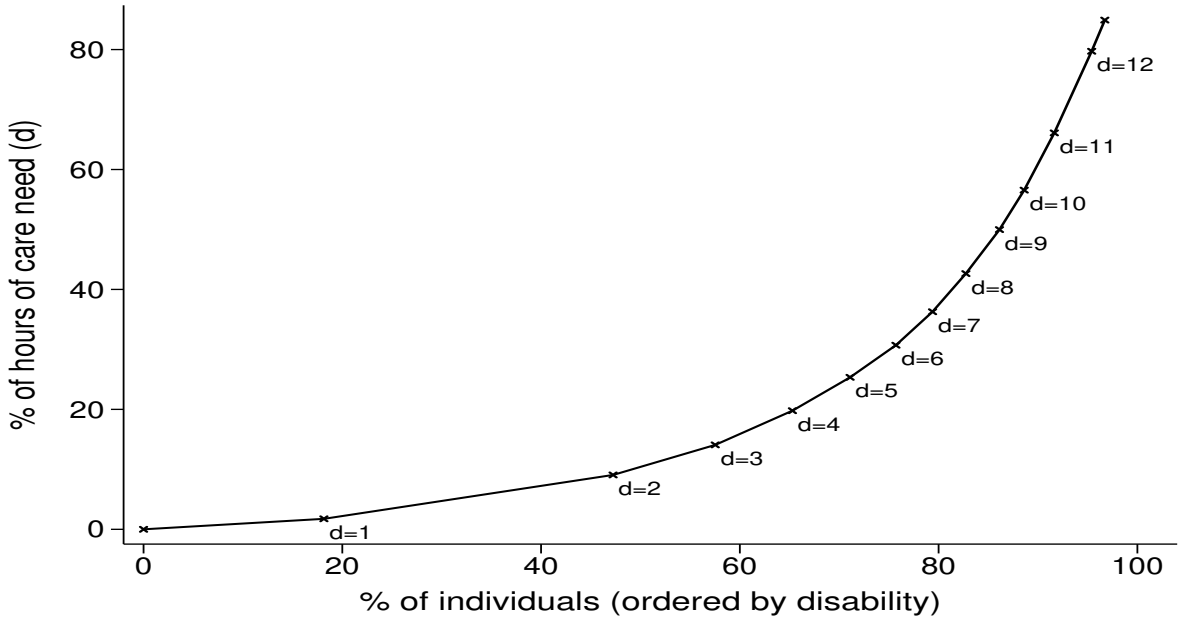
Sample: All respondents aged 65 and above who received care. SHARE: Waves 1-2. HRS: Waves 5-11. Weights are used. Bins for care need d in daily hours: (0,2], (2-4], ..., > 12.

NH, see Figure 13 in the appendix). Overall, among singles FHC increases in importance, but it remains below IC even in the North and the Middle, where public policy places a considerable emphasis on it. Only in Denmark does FHC surpass IC (for singles).

But who are the informal caregivers? Figure 9 splits up IC hours into where they come from: from a working-age caregiver in the family (*young*, mostly children of the elderly), a retirement-age caregiver in the family (*old*, mostly spouses), or other friends and relatives (*other*) for whom age is unknown. We see that in all regions, single elderly receive care mainly from young caregivers. Among these, we find that it is mostly daughters who take on the task of caregiving (more on this below). As for coupled elderly, almost all care is given by old caregivers; these are almost always spouses.

Result 7 *The coupled elderly are predominantly cared for by their spouse at home in the Middle, South, and U.S., but in the North formal care is as important as care from the spouse. Among singles, there is more variation in care arrangements: nursing homes are most prevalent in North and Middle, whereas IC is most common in South and remains very significant in the U.S.*

FIGURE 7: Lorenz Curve of Care



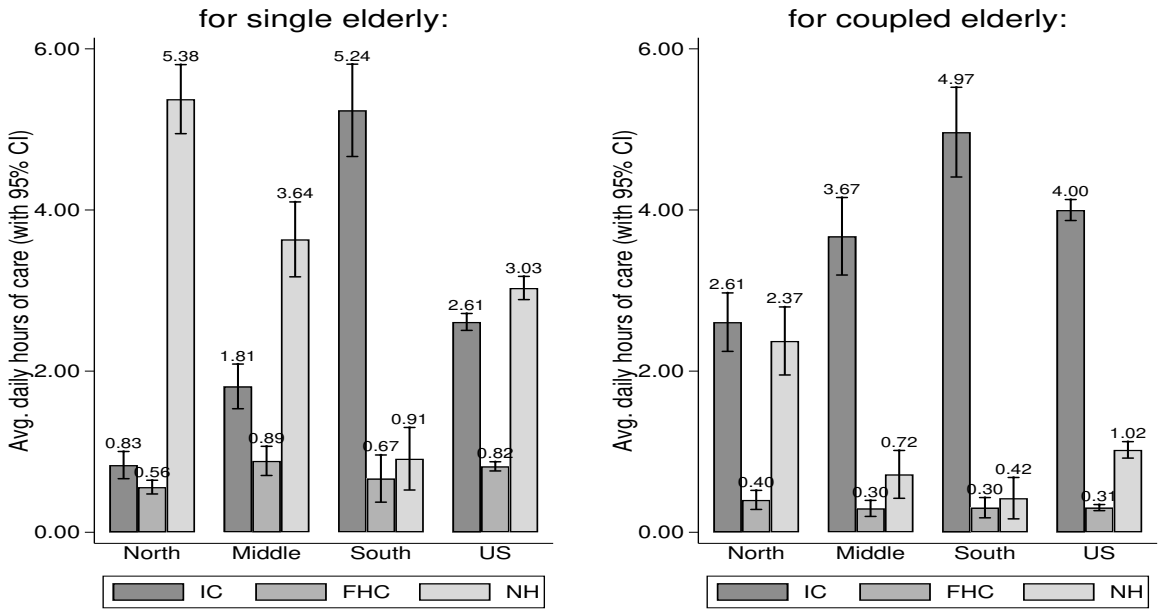
For integer values of d (point markers), graph shows % of individuals with care need of d and below on the x-axis and the fraction of total care needs in the population as measured by d on the y-axis. We use care need and not reported care hours since care hours are not reported by NH residents. Sample: All respondents aged 65 and above with limitations ($d > 0$), all waves from SHARE and HRS. Weights are used. Labels give disability score pertaining to point on the curve.

IV.3.d. The importance of co-resident caregivers

We now zoom further into our sample and consider only those respondents who received IC. An important distinction among the IC population is whether a caregiver co-resides with the elderly in need or not. Especially for the most dependent individuals, it is often necessary that a caregiver is at reach around the clock. Thus, we expect a large number of hours to come from co-residing caregivers. Figure 10 splits all IC hours into those provided by a co-resident (split again into partners and children – others are negligible) and those provided by caregivers from outside the household (most often children). Indeed, the following fact emerges:

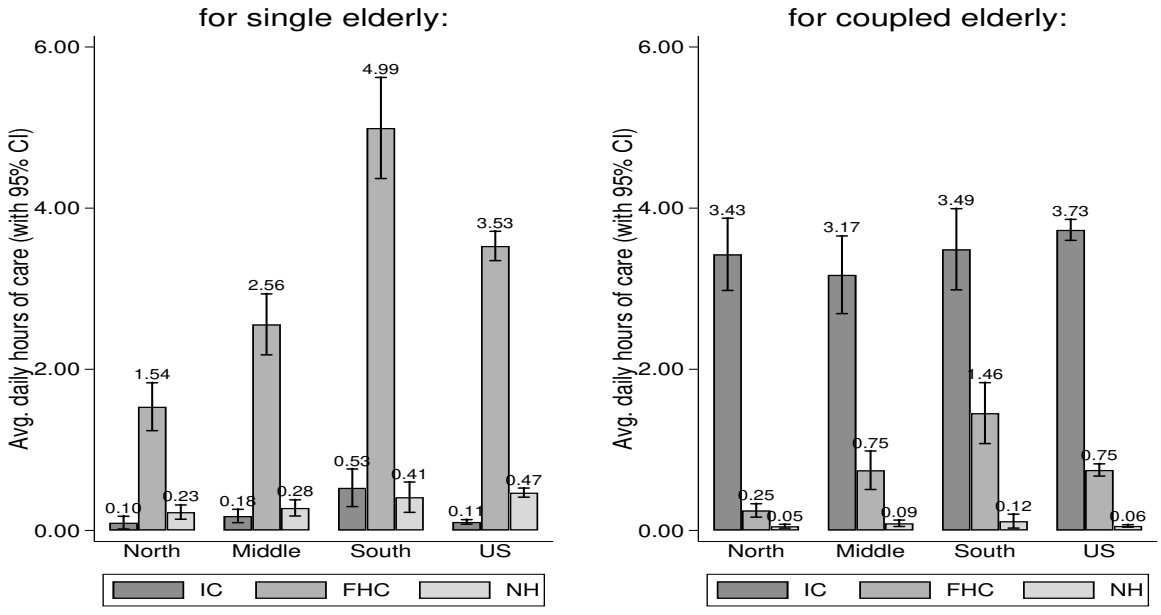
Result 8 *Co-resident caregivers give about two-thirds of all IC hours, in all regions. Even for single elderly, co-resident caregivers are as important as outside-household caregivers in the South and, to a lesser extent, in the U.S.*

FIGURE 8: Care choice by coupledness status



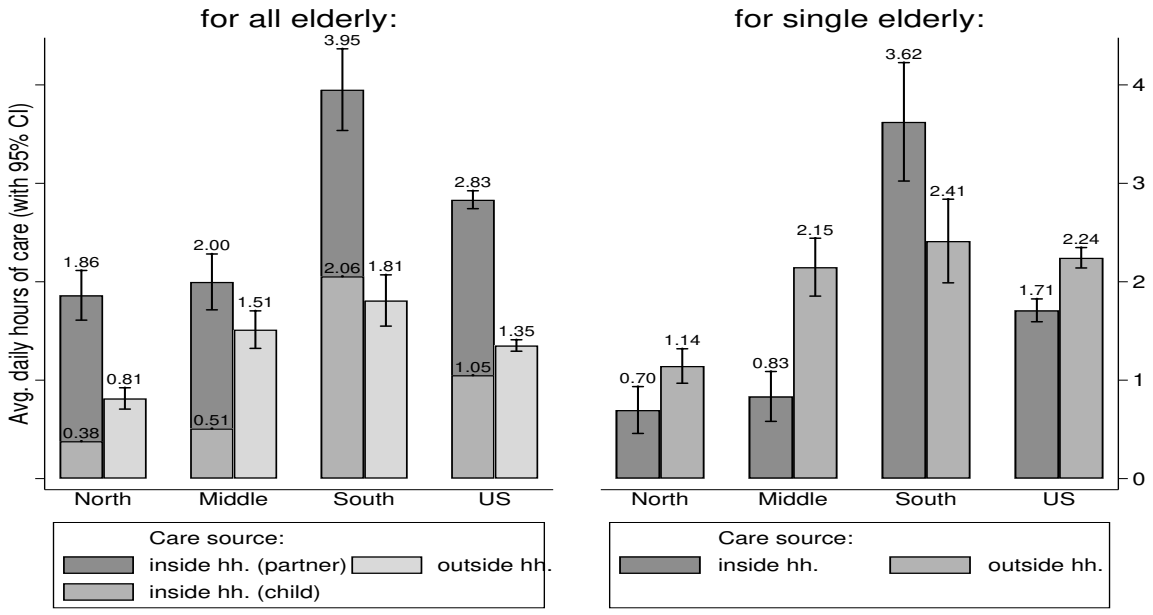
Sample: all respondents who received care. SHARE: Waves 1-2. HRS: Waves 5-11. Weights are used.

FIGURE 9: Source of informal-care hours



Sample: All respondents who received IC. SHARE: Waves 1-2, HRS: Waves 5-11. Weights are used.

FIGURE 10: Inside-household versus outside-household informal care



Sample: All respondents who received IC. SHARE: Waves 1-2, HRS: Waves 5-11. Weights are used.

IV.3.e. The main providers of IC

We now shift our attention to the providers of IC. As we have seen, IC hours can be very substantial, amounting to a part-time job and in some cases even to a full-time job. However, if the care burden were shared among many family members, this need not imply a large individual burden on each caregiver. But we find that the opposite is true: Indeed, care provision is usually concentrated on *one* person. We calculate the share of hours that the main informal caregiver (the one with most hours) contributes to all IC hours that a person receives. Figure 11 shows the distribution of this share for those elderly with above 3 hours of care needs per day who receive their care mainly from informal sources. We see that even for singles, for which one might conjecture that the children share the caregiving burden, it is usually one person who contributes more than 80% of the IC hours (the right-most bar of the histograms). In less than 20% of cases (and only in 10% of cases for U.S. and the South), the main caregiver contributes 60% or less. For the coupled, concentration of care hours on the spouse is even more extreme.

Finally we ask who the persons are who provide the lion's share of IC. We define a *heavy helper* (HH) as an informal caregiver who gives 3 or more daily hours of care (amounting to 21 hours or more per week, i.e. a part-time job). Table 8 shows characteristics of heavy helpers and compares them to other (comparable) children.

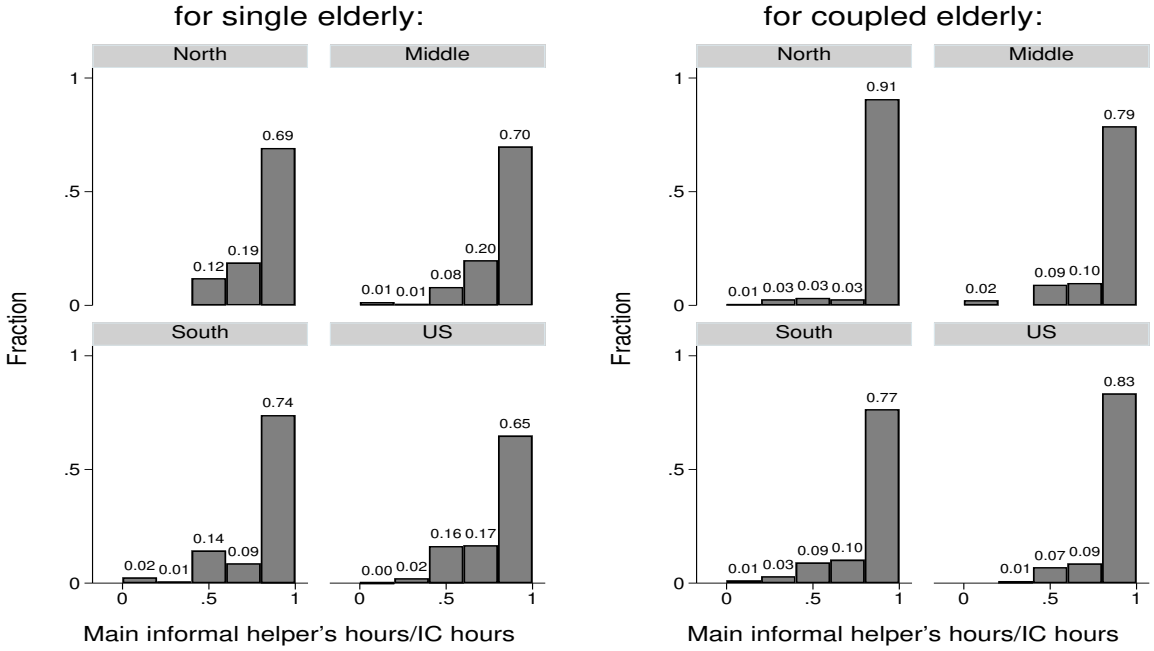
Heavy-helping (HH) spouses are more likely to be female in all regions. This reflects two facts:

TABLE 8: Characteristics of heavy-helping (HH) spouses, HH kids and comparable non-HH kids

Variable	Group	North	Middle	South	U.S.
female	HH spouses	0.52	0.70	0.62	0.54
	HH kids	0.85	0.70	0.84	0.74
	Non-HH kids	0.51	0.52	0.49	0.50
average age	HH spouses	75.2	74.9	74.1	68.5
	HH kids	56.1	55.0	54.0	48.6
	Non-HH kids	56.3	56.4	55.0	52.4
co-residing	HH kids	0.61	0.35	0.64	0.62
	Non-HH kids	0.03	0.06	0.16	0.06
same building	HH kids	0.64	0.53	0.72	–
	Non-HH kids	0.03	0.12	0.23	–
< 5 km	HH kids	0.88	0.80	0.94	0.67
	Non-HH kids	0.40	0.38	0.61	0.33
married	HH kids	0.51	0.46	0.76	0.45
	Non-HH kids	0.70	0.63	0.76	0.69
# children	HH kids	1.59	1.52	1.45	1.92
	Non-HH kids	1.80	1.65	1.70	2.23
≥ 65 years	HH kids	0.22	0.16	0.13	0.10
	Non-HH kids	0.20	0.20	0.18	0.16
≥ 62 years	HH kids	0.26	0.26	0.21	0.16
	Non-HH kids	0.31	0.32	0.28	0.25
Full-time	HH kids	0.23	0.23	0.25	0.39
	Non-HH kids	0.44	0.41	0.43	0.63
Part-time	HH kids	0.14	0.19	0.07	0.12
	Non-HH kids	0.10	0.08	0.03	0.09
Self-empl.	HH kids	0.13	0.08	0.03	–
	Non-HH kids	0.07	0.09	0.05	–
retired	HH kids	0.29	0.23	0.16	–
	Non-HH kids	0.23	0.29	0.16	–
inactive	HH kids	0.21	0.28	0.50	0.48
	Non-HH kids	0.17	0.14	0.32	0.28
educ. Yrs.	HH kids	12.11	11.25	8.80	12.8
	Non-HH kids	11.62	11.61	8.81	13.2
N	HH spouses	319	202	239	2,912
	HH kids	47	92	227	2,951
	Non-HH kids	3,110	2,490	2,941	9,011

Table shows means of female dummy, age, and a coresidence dummy for spouses and children who are heavy helpers (primary informal caregiver who provides 3 or more hours of daily care). SHARE: Waves 1, 2, 5, and 6 (only use helpers for whom hours can be determined in Waves 5-6.). HRS: Waves 5-11. Weights are used.

FIGURE 11: Concentration of informal caregiving on one caregiver



Sample: All respondents aged 65 and above who are disabled ($d \geq 3$) and receive more than half of their hours of care from informal sources. SHARE: Waves 1-2; HRS: Waves 5-11. Weights are used.

First, women tend to become dependent and die later than men. Second, women tend to be married to older men. The two together make it likely that the husband experiences dependency earlier and thus the wife is more likely to care for her spouse than the other way around. HH spouses are usually in retirement age, such that is little conflict with labor-market activities.

Among HH *children*, an even higher fraction is female: About three-quarters or more of heavy helpers are daughters. HH spouses and children in the U.S. are substantially younger than their counterparts in Europe, presumably reflecting the earlier onset of severe dependency and earlier child-bearing in the U.S.

As may be expected, physical proximity between HH children and their frail parents plays an important role. Co-residence rates are a lot higher for HH children than for other children. In the South, almost all HH children live within a few kilometers of their parents. Among non-HH children, this is not the case, much less so in North, Middle, and the U.S. High co-residence rates between HH children and their parents highlight an interaction that presumably runs both ways: Intense caregiving requires physical proximity, and being physically close makes a child more likely to become a heavy helper.

HH children appear to have slightly fewer competing time demands from their own families.

They are less likely to be married and have fewer children than comparable non-HH children.

In terms of labor-market characteristics, we observe that the vast majority of HH children are still of working age and so caregiving bears potentially high opportunity costs in terms of forgone labor income. HH children hold full-time jobs less often than non-HH children, especially in the U.S. However, there is still a quite large number of HH children who are working alongside their heavy caregiving activity, again especially so in the U.S.

Result 9 *Heavy-helping children tend to be female and of working age. They live close to their parents or co-reside and are less active in the labor market than comparable children.*

V. Robustness of results

All results for care hours so far relied on our benchmark imputation scheme (based on the Dahl selection model). The reader may wonder how robust our results are with respect to this choice. In Table 9, we show how our main result from Fig. 4 changes when using alternative imputation schemes for hours. As we already saw in the horse race between the models, imputed hours for IC and FHC are very similar across model specifications, showing that the results are very robust for these forms of care. The main difference between the models arises in the imputation of community-equivalent NH hours, where our baseline model corrects for the selection bias and ascribes substantially higher numbers than OLS in all countries; however, this correction is of similar magnitude for all countries (proportionally, at least), thus the implied cross-country gradients are again very robust. Finally, Table 9 shows in the third-to-last column that our results barely change when omitting couples for whom the assignment of household-level care hours is problematic; thus the results are robust also with respect to Ass. 2 (on how to assign household hours).

VI. Conclusions

We have proposed a methodology of how to overcome several challenges in constructing comparable LTC data across Europe and the U.S. At the center of our approach is a selection model into different care forms that allows us to impute community-equivalent care hours for missing hours observations. We have then used the constructed data to quantify the landscape of all care sources within and across countries. We found informal care to be most important in southern European countries. In the U.S., informal care is more important than formal care and exceeds informal care given in central European countries, where informal care and nursing homes are of similar importance. Nursing homes are the preferred care form in northern European countries, where governments play a larger role in LTC than in the other countries in our sample. Formal-home care

TABLE 9: Robustness checks for main result

country	source of care	Dahl selection	elim. probl. cpls	grouped OLS	pooled OLS
Netherlands	IC	0.15	0.14	0.13	0.12
	FHC	0.07	0.07	0.07	0.07
	NH	0.86	0.86	0.53	0.53
Sweden	IC	0.26	0.24	0.20	0.18
	FHC	0.10	0.10	0.11	0.11
	NH	0.95	0.95	0.58	0.58
Denmark	IC	0.22	0.18	0.17	0.15
	FHC	0.14	0.13	0.15	0.15
	NH	0.67	0.67	0.41	0.41
Belgium	IC	0.46	0.41	0.40	0.35
	FHC	0.12	0.11	0.12	0.12
	NH	0.87	0.87	0.48	0.48
NORTH	IC	0.27	0.25	0.23	0.20
	FHC	0.10	0.09	0.10	0.10
	NH	0.86	0.86	0.51	0.51
France	IC	0.49	0.46	0.41	0.37
	FHC	0.15	0.15	0.16	0.16
	NH	0.57	0.57	0.32	0.32
Germany	IC	0.46	0.43	0.48	0.43
	FHC	0.11	0.11	0.11	0.11
	NH	0.48	0.48	0.25	0.25
Austria	IC	0.52	0.49	0.56	0.48
	FHC	0.18	0.16	0.18	0.17
	NH	0.25	0.25	0.13	0.13
MIDDLE	IC	0.47	0.44	0.46	0.41
	FHC	0.13	0.13	0.13	0.13
	NH	0.50	0.50	0.27	0.27
Spain	IC	0.97	0.94	0.93	0.80
	FHC	0.07	0.07	0.07	0.07
	NH	0.24	0.24	0.13	0.13
Italy	IC	0.77	0.72	0.77	0.66
	FHC	0.10	0.10	0.10	0.09
	NH	0.04	0.04	0.02	0.02
SOUTH	IC	0.85	0.81	0.84	0.72
	FHC	0.09	0.09	0.08	0.08
	NH	0.12	0.12	0.07	0.07
US	IC	0.61	0.61	0.61	0.61
	FHC	0.12	0.12	0.12	0.12
	NH	0.40	0.40	0.25	0.25

The table shows average hours of care received by all respondents aged 65 and above calculated under various assumptions. The column *Dahl selection* is our benchmark model and replicates the average hours shown in Figure 4. The column *elim. probl. cpls* we use the baseline model leaving out couples in SHARE for which household-level and not individual-level care hours are known. The final two columns use imputations from the grouped-OLS and pooled-OLS models.

plays a quantitatively fairly moderate role in all countries, which is to say that the lion's share of care hours in the community is borne by family members. Within the family, we find that the main care burden is on one caregiver.

Our results are suggestive that policy influences care choices in important ways. However, cultural factors also affect these choices and are certainly reflected in existing policies. Future work should therefore aim to disentangle the extent to which policy relative to societal values matter in shaping care choices. Understanding the quantitative significance of these channels would be useful for evaluating the costs and welfare effects of LTC-reform scenarios in different countries.

Acknowledgments

We thank the editor Monica Costa Dias and anonymous referees for useful suggestions. We thank David San-Eloy for excellent research assistance. We thank conference participants at the SHARE workshop in Munich (2016) and at the SAE Simposio (Bilbao, 2016) as well as seminar participants at U. Autonoma de Barcelona and Carlos III for comments. Daniel Barczyk acknowledges research funding through the SSHRC Insight Grant (435-2018-0754) and grant FRQ-SC NP-191888. Matthias Kredler acknowledges research funding by the Fundación Ramón Areces, by the Spanish Ministerio de Economía y Competitividad, grants ECO2012-34581, ECO2015-68615-P, and MDM 2014-0431, and Comunidad de Madrid, MadEco-CM (S2015/HUM-3444).. M. Kredler gratefully acknowledges research funding by *Fundación Ramón Areces*; the *Ministerio de Economía y Competitividad* (Spain), grants MDM 2014-0431 and ECO2015-68615-P; and Comunidad de Madrid, MadEco-CM (S2015/HUM-3444). This paper uses data from SHARE Waves 1, 2, 3 (SHARELIFE), 4, 5 and 6 (DOIs: 10.6103/SHARE.w1.600, 10.6103/SHARE.w2.600, 10.6103/SHARE.w3.600, 10.6103/SHARE.w4.600, 10.6103/SHARE.w5.600, 10.6103/SHARE.w6.600), see BÃ¶rsch-Supan et al. (2013) for methodological details. The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: NÂ°211909, SHARE-LEAP: NÂ°227822, SHARE M4: NÂ°261982). Additional funding from the German Ministry of Education and Research, the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, HHSN271201300071C) and from various national funding sources is gratefully acknowledged (see www.share-project.org).

References

Bakx, P., de Meijer, C., Schut, F. & van Doorslaer, E. (2015), 'Going formal or informal, who cares? the influence of public long-term care insurance', *Health Economics* (24), 631–643.

Balia, S. & Brau, R. (2014), 'A country for old men? Long-term home care utilization in Europe', *Health Economics* **23**(10), 1185–1212.

Barczyk, D. & Kredler, M. (2018), 'Evaluating long-term-care policy options, taking the family seriously', *Review of Economic Studies* **85**(2), 766–809.

Bolin, K., Lindgren, B. & Lundborg, P. (2008*a*), 'Informal and formal care among single-living elderly in Europe', *Health Economics* **17**(3), 393–409.

Bolin, K., Lindgren, B. & Lundborg, P. (2008*b*), 'Your next of kin or your own career?: Caring and working among the 50+ of Europe', *Journal of Health Economics* **27**(3), 718–738.

Bonsang, E. (2009), 'Does informal care from children to their elderly parents substitute for formal care in Europe?', *Journal of Health Economics* **28**(1), 143–154.

Börsch-Supan, A. & Jürges, H., eds (2005), *The Survey of Health, Ageing and Retirement in Europe – Methodology*, Mannheim Research Institute for the Economics of Aging (MEA).

Bourguignon, F., Fournier, M. & Gurgand, M. (2007), 'Selection bias corrections based on the multinomial logit model: Monte carlo comparisons', *Journal of Economic Surveys* **21**(1), 174–205.

Crespo, L. & Mira, P. (2014), 'Caregiving to elderly parents and employment status of European mature women', *Review of Economics and Statistics* **96**(4), 693–709.

Dahl, G. B. (2002), 'Mobility and the return to education: Testing a roy model with multiple markets', *Econometrica* **70**(2), 2367–2420.

Dobrescu, L. I. (2015), 'To love or to pay: Savings and health care in older age', *Journal of Human Resources* **50**(1), 254–299.

Kars, M. C., van Thiel, G. J., van der Graaf, R., Moors, M., de Graeff, A. & van Delden, J. J. (2016), 'A systematic review of reasons for gatekeeping in palliative care research', *Palliative medicine* **30**(6), 533–548.

Lipszyc, B., Sail, E., Xavier, A. et al. (2012), Long-term care: need, use and expenditure in the EU-27, Technical report, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Malter, F. & Börsch-Supan, A., eds (2013), *SHARE Wave 4: Innovations & Methodology*, Munich Center for the Economics of Aging.

OECD (2017), *Health at a Glance 2017: OECD Indicators*, OECD Publishing, Paris.

OECD/EU (2013), *A good life in old age? Monitoring and improving quality in long-term care*, OECD Publishing, Paris.

United Nations (2017), 'World population prospects: The 2017 revision'.
URL: *https://esa.un.org/unpd/wpp/publications/*

A Appendix

I.1. Sampling, retention, and weights

When SHARE first contacts households the response rate is between 43% to 68%, varying by country. Inside each household, 89% of eligible individuals in the households respond. Like any panel, SHARE suffers from attrition through mortality and respondents withdrawal. The retention rates of alive individuals in SHARE are lower than in the HRS. Malter & Börsch-Supan (2013) document in their Chapter 10 that the probability of household response in SHARE's Wave 4 conditional on response in Wave 3 was around 80%, whereas it is above 90% for the HRS. In the long run, the retention rates for households first contacted in Wave 1 and who are still in Wave 4 drops to around 70%. Between countries, it varies from below 60% (Austria, Germany, Netherlands) to 90% (Switzerland). Retention at the *individual* level, however, is unproblematic: Conditional on a household being retained, all individuals in the household are almost certain to respond again. When a person dies, SHARE attempts to locate a proxy respondent (typically a close relative of the deceased) to carry out an *end-of-life interview* in order to learn about the circumstances of the deceased at and before death.

SHARE provides sampling-design weights to compensate for unequal selection probabilities of households/individuals and for non-response of sample units; whenever appropriate, we use the "calibrated cross-sectional household weights" (`cchw_w#`) in order to obtain unbiased population statistics, following the recommendation of Malter & Börsch-Supan (2013) (see Ch. 8). For the HRS, we use the weights provided by RAND (`RwWTCRNH`) which include the nursing-home weights.

I.2. Evolution of nursing-home population in SHARE and HRS

Table 10 shows the evolution of the nursing-home population in SHARE and the HRS across waves and compares it to the numbers provided by the OECD.

TABLE 10: Percentage of over-65-year-olds in nursing homes over time

Country	Source	2004	2007	2009	2011	2013	2015
Sweden	SHARE	1.9	4.2	3.4	2.8	2	2.6
	OECD	6.8	6	5.8	5.2	4.9	4.5
Netherlands	SHARE	4.1	4.4	3.5	3	4.2	–
	OECD	7.3	6.9	6.7	6.5	5.6	5.3
Denmark	SHARE	4.2	4.2	3.6	3.4	3.7	2.9
	OECD	–	5.1	4.7	4.4	4	3.9
Belgium	SHARE	0.3	1.8	2.9	4	4	4.3
	OECD	6.4	6.6	6.7	6.7	8.8	8.8
Austria	SHARE	1.3	1.6	2.1	2.9	2.3	2.2
	OECD	–	–	–	–	–	–
Germany	SHARE	1.3	2.5	2.3	2	2.4	1.6
	OECD	3.8	3.8	3.8	4	4.1	4.1
France	SHARE	0.7	1.8	1.9	1.5	1.7	2.3
	OECD	3.7	4	4.2	4.4	4.4	4.2
Spain	SHARE	0.6	1.2	1.4	1.2	1.5	1.3
	OECD	–	–	1.3	1.7	1.8	1.8
Italy	SHARE	0.5	0.3	0.1	0.5	0.9	1.1
	OECD	–	–	–	–	–	–
U.S.	HRS	4.1	4.0	3.5	3.1	2.9	
	OECD	3.9	3.7	3.4	3.3	–	–

OECD data are from `http://stats.oecd.org`, Section *Long-Term Care Resources and Utilisation*; if data were not available for a year in question we took data (i) from the subsequent year and, if that was not available, (ii) from the year before; if neither (i) nor (ii) was available we report "–" (no data available). SHARE and HRS statistics are computed using the nursing-home dummy and weights.

I.3. Excluding help to healthy respondents

In general, we set care dummies and care hours to zero for all care categories if respondents did not report any (I)ADL limitations. We do this since LTC is defined as help with (I)ADLs, so any positive care hours reported by respondents without limitations are not consistent with this definition. Respondents sometimes seemed to declare help by cleaning ladies, home aides, and regular help by family members as care; we want to exclude such help from our measure of care.

I.4. Nursing-home dummy

This dummy is 1 if either (i) the respondent was interviewed in a nursing-home or (ii) the respondent was interviewed at home but stated that (s)he was a permanent resident of a nursing

home.[34] For the regular waves (1, 2, 4, 5), we take the information about (i) from the dummy `nursinghome` that SHARE provides in their imputations module. Case (ii) is identified from question `hc029` in the health-care module.[35] In Wave 3, we inferred (i) from questions `iv012` (interviewer stated interview took place in "special housing for the elderly (24 hours attention)") and `iv021` (interviewer stated that the proxy who answered in lieu of the respondent was nursing-home staff) and (ii) from question `ac011` (the respondent gave a nursing home as the current residence when answering on the history over all housing units/accommodations over his/her life-time.[36]

I.5. Informal care (IC)

In all waves but Wave 3, SHARE asked about IC. In Waves 1,2 and 5, all information about IC is contained in the Social-Support Module. Waves 4 and 6 are special: We had to merge the Social-Support Module with the Social-Networks Module in order to determine the identity of some caregivers (see questions `sp003` and `sp021` as described in Release 1.1.1 of Wave 4, p. 11-13 and Release 6.0.0 of Wave 6, p. 28-30.) Unlike the HRS, SHARE has different sets of questions on IC coming from caregivers who live *inside* the respondent's household (often spouses) and caregivers who live *outside* the respondent's household (often children).

1. *IC from persons outside the household (outside-household informal care, OIC):* Inside each household, SHARE asks one person (the so-called *family respondent*) if anybody in the household is receiving help from family members or friends who live outside the household (questions `sp002` and following) and who these helpers are. In Waves 1 and 2, there is also a question on how frequently such help was received from each helper (`sp005`: daily, weekly, monthly, or less), and how many hours of care were received on a typical day/week/month/in the last year (`sp006`). However, Waves 4, 5 and 6 do not contain information on hours but only on frequency (daily, weekly, monthly, or less). What complicates matters further is that the question on hours received in Waves 1 and 2 asks about the hours of care received *altogether by all members of the household*. Since we are after hours at the individual and not the household level, we proceed as follows to split up hours between couples:[37]

[34] It may also be that a proxy respondent, often the respondent's spouse, answered the questionnaire for the respondent and stated that the respondent was in a nursing home.

[35] Note that the SHARE dummy `nursinghome` coincides with our nursing-home dummy in Wave 1; in Wave 1, case (i) does not occur yet and SHARE codes case (ii) as a 1. In Waves 2, 4, and 5, however, the two differ because SHARE does not code case (ii) as 1 any more: Our nursing-home-resident dummy includes cases (i) and (ii), whereas the SHARE dummy only includes case (i).

[36] The nursing-home dummy as constructed in Wave 3 seems consistent with the dummy from the other waves. We found no breaks in the time series of nursing-home residents for any country. Also the nursing-home entry and exit rates look stationary.

[37] Although it is not foreseen by the SHARE manuals, the question on informal care from persons outside the

(a) *one partner in nursing home (less than 1% of cases)*: If one member of the couple resides in a nursing home, we assign all OIC hours to the household member at home and none to the nursing-home resident.

(b) *only one partner dependent (about 75% of cases):* If one of the partners reported at least one (I)ADL limitation but the other partner reported zero limitations (or the (I)ADL information was missing), we assign all OIC hours of care to the household member with limitations and zero to the one without limitations.

(c) *both partners dependent (remaining cases):* If both partners had at least one (I)ADL limitation, then we split the OIC hours between them. The split is proportional to a weighted sum of their (I)ADL limitations; we weigh the (I)ADLs since we find that each of them necessitates different amounts of care. The weights are taken from a regression of daily hours of care received by HRS respondents (for which we know care hours on an individual basis) on the (I)ADL dummies and a dementia dummy, omitting the constant term.

2. *IC from persons inside the household (inside-household informal care, IIC):* In households consisting of more than one person, SHARE asks each survey-eligible person both if she or he (i) gave personal care to or (ii) received personal care from other persons in the same household. *Personal care* is defined by the question as help with ADLs such as washing, dressing, getting out of bed that occurred on a daily or almost daily basis. A problem we encountered with inside-household help was that within couples the answers on who gave care to whom was often contradictory. For example, in some cases the husband reported to have received care from his wife, but the wife said she did not give help to her husband. This might be due to true disagreement between the partners; for example, the dependent partner may not want to admit that he was being helped. However, our analysis of the data rather suggests that these inconsistencies are due to data-entry errors. Consistent with this hypothesis, we find that disagreement between the partners' answers is a lot more likely when a proxy answered the questions about caregiving for the respondent. This proxy is almost always the spouse him/herself – in this case, there is obviously no room for disagreement, since one person answered for both partners. We conclude that most likely the interviewer and respondent were not careful when going over the questions about caregiving to the partner a second time (in the case of proxy interviews) or when spouses answered the survey together and had to answer the same questions from the perspective of both sides. We adopt the following strategy to address this issue: Whenever there is contradictory information inside a couple

household was often answered separately by inhabitants of the same household. The separate answers look plausible if interpreted as referring to care that the respondent received him/herself; in such cases we thus determine care hours separately from each partner's answer.

or when the partners claimed that they were giving care to each other, we check if both members of the couple are actually dependent. We assume (i) that non-dependent persons do not receive care, (ii) that couples with a similar level of dependency really provide care to each other (they may have problems with different activities, making this possible), and (iii) that care goes from the healthier to the more dependent person in case of large differences in dependency. As a *large difference*, we define a difference of 3 or higher in the count of the 6 ADL limitations (note that the question explicitly referred to help with these ADLs).

I.6. Formal home care (FHC)

SHARE Waves 1, 2, 5 and 6 collected information on formal home care (FHC); Waves 3 and 4 did not ask about FHC. Questions on FHC were asked on the *individual* level, meaning that we have separate information on each household member. SHARE divides the question on FHC into three categories: (i) *nursing/personal care*, (ii) *private domestic help*, and (iii) *meals-on-wheels*. Nursing/personal care explicitly refers to activities that fall strictly under LTC: help with dressing, bathing, showering, going to bed etc. Private domestic help includes typical household tasks like cleaning, cooking, and shopping.[38] Meals-on-wheels refers to ready-made meals provided by an organization or company to the respondent. In Waves 1 and 2, SHARE asks precisely for how many weeks in the last year and how many hours per week each of the three types of formal home care was received (however, hours are not asked for when it comes to meals-on-wheels). In Waves 5 and 6, there is only information *if* the respondent received each type of care, but not on the frequency nor on hours. Whenever the information is available, we compute the hours of care for each type of FHC as follows. We multiply the number of weeks in which care was received in the last year (questions `hc033`, `hc035`, `hc037`) by the number of weekly hours (questions `hc034`, `hc035`). For meals-on-wheels, there is no information on weekly hours; we impute 2.12 hours per day. We obtain this number from the following regression: We regress the total daily hours of care received on 11 dummies for the (I)ADLs and a dummy for memory limitations, omitting the constant term. The sample for this regression are non-nursing-home residents in the HRS. The coefficient on the IADL dummy *preparing meals* is 2.12, which has the interpretation as the daily time need of preparing meals for a caregiver at home.

[38]The explanatory text to the question on paid domestic help makes it clear (in all waves) that the question refers only to help that was received because the respondent had health problems that made the task difficult to perform for the respondent herself. Thus, in theory, cleaning ladies hired by perfectly healthy respondents should not be reported here. However, the data indicate that the question was not always understood this way by respondents. Out of the roughly 6,200 respondents who said they received help with domestic tasks, about 750 reported that they had no mobility limitations. We did not count help to these apparently healthy respondents as care.

I.7. Imputation of total care hours

This appendix gives details on the imputation of total care hours for cases in which they are missing. We first describe the pooled and grouped OLS methods. This will facilitate the discussion of the benchmark imputation method, the Dahl selection model.

Pooled OLS. We run a OLS regression of total care hours on IADL-limitation dummies, a memory-limitation dummy, and age. The sample consists of HRS and SHARE community residents for whom we have precise information on total care hours. Note that in SHARE, these can only be respondents who received OIC and/or FHC in Waves 1 and 2. We use predicted values from this regression to impute total care hours in all cases where they are unknown.

Grouped OLS. We estimate regressions using OLS differentiating between two samples of community residents: In one, we include all those respondents who receive (at least some) IIC and the other consist of individuals who receive no IIC (i.e. those who receive exclusively FHC and/or OIC). We do so since having an inside-household caregiver is likely to contain information on care needs that go beyond the observables in our data. The additional information content from IIC is that co-residing with a helper is often associated with a helper being needed around the clock; Barczyk & Kredler (2018) document that co-residency between children and parents is frequently associated with heavy care needs. In all regressions, we use the same right-hand side variables as in pooled OLS: IADL-limitation dummies, the memory-limitation dummy, and age.

Specifically, we impute hours as follows for the following three groups:

1. *Non-IIC community residents with missing hours* (more precisely: community residents in SHARE and HRS who do not receive IIC and for whom we know *that* they received OIC and/or FHC, but for whom information is lacking on *hours* of OIC and/or *hours* of FHC; most of these cases come from Waves 5 and 6 in SHARE, in which hours were not asked for): We impute total hours for these individuals from the sample of HRS and SHARE community residents who do not receive any care from inside the household and for whom total hours are known.

2. *IIC recipients in SHARE* (more precisely: for all SHARE respondents who received some IIC, but also HRS community residents who received some IIC but for whom hours information is incomplete): We impute total hours of care for these individuals from the sample of HRS community residents who receive care from at least one member of the same household and for whom we have precise hours information.

3. *Nursing-home residents in SHARE and HRS*: Since we do not know if NH residents would have been given IIC or not if they were in the community, we impute total hours for this

group based on a regression for the entire community sample, meaning that imputations for this group are identical to the pooled-OLS scheme.

Dahl selection model. The main text describes how we obtain the estimate for the population mean of hours, $x'_{it}\beta$, for each dependency count. To infer expected hours of care of NH and community residents in each country, we then use the following identity:

$$\mathbb{E}[\epsilon_{it}|x_{it}] = \mathbb{P}(NH|x_{it}, z_{it})\mathbb{E}[\epsilon_{it}|x_{it}, z_{it}, NH] + \mathbb{P}(C|x_{it}, z_{it})\mathbb{E}[\epsilon_{it}|x_{it}, z_{it}, C] = 0, \tag{5}$$

where C stands for being in the community and $\mathbb{P}$ stands for the probability of an event. We now make use of the fact that $\mathbb{E}[\epsilon_{it}|P_{it}] = \mu(P_{it})$ and use the probit estimates $\hat{P}_{it}$ and $1 - \hat{P}_{it}$ for $\mathbb{P}(NH|x_{it}, z_{it})$ and $\mathbb{P}(C|x_{it}, z_{it})$, respectively, to obtain estimates for hours of community and NH residents:

$$\hat{h}^{C}_{it} = x'_{it}\hat{\beta} + \hat{\mu}(\hat{P}_{it}), \tag{6}$$

$$\hat{h}^{NH}_{it} = x'_{it}\hat{\beta} - \frac{1 - \hat{P}_{it}}{\hat{P}_{it}}\hat{\mu}(\hat{P}_{it}). \tag{7}$$

We use Eq. (7) to impute hours of NH residents (in both the HRS and SHARE).

The first three columns of Table 11 show how the Dahl model corrects for selection, comparing it to pooled OLS, reporting hours means at the country level. As we see in the first three data columns of the table, the correction plays out as expected. The estimator infers that NH residents are negatively selected, thus community hours are about five to six hours below NH hours in all countries, the population mean lying between the two. Compared to the pooled-OLS imputations, we see that the Dahl corrections are highest in the South (above seven hours), they are medium-size (about six hours) in the U.S. and Middle, and they are lowest (4.5 to 6 hours) in the North. This is consistent with Southern countries having the lowest NH populations, thus selection being only working on the frailest individuals, whereas the NH population is a larger subset of the disabled the farther we move North. We note that selection correction thus attenuates the North-South gradient of NH hours somewhat with respect to pooled OLS.

For the SHARE sample, we need an additional step to infer hours for community residents who receive care from somebody within the household (IIC). Recall that among community residents in SHARE's Waves 1 and 2, we only observe hours for the subsample of respondents who receive care exclusively from sources outside the household (OC). The following identity holds:

$$\mathbb{E}[h_{it}|x_{it}, C] = \mathbb{P}(OC|x_{it}, z_{it}, C)\mathbb{E}[h_{it}|x_{it}, z_{it}, OC] + \mathbb{P}(IIC|x_{it}, z_{it}, C)\mathbb{E}[h_{it}|x_{it}, z_{it}, IIC]. \tag{8}$$

In order to impute IIC hours, we again replace population variables by their finite-sample analogues

TABLE 11: Imputations for NH hours

Country	$x'_{it}\hat{\beta}$	$\hat{h}^C$	$\hat{h}^{NH}$	Pooled OLS	**Dahl correction**
Netherlands	8.65	6.61	12.07	7.40	**+4.67**
Sweden	10.97	9.31	14.78	9.09	**+5.69**
Denmark	9.00	7.30	12.77	7.82	**+4.95**
Belgium	9.00	7.95	13.41	7.41	**+6.00**
France	10.06	9.30	14.76	8.25	**+6.51**
Germany	7.93	7.13	12.60	6.65	**+5.95**
Austria	7.77	7.13	12.60	6.47	**+6.13**
Spain	13.61	13.06	18.52	10.05	**+8.47**
Italy	9.90	9.58	15.05	7.84	**+7.21**
US	13.07	10.85	16.31	10.18	**+6.13**

Table shows means over countries for all waves regarding NH imputations.

to obtain

$$\hat{h}_{it}^{IIC} = \frac{\hat{h}_{it}^{C}}{\hat{P}_{it}^{IIC|C}} - \frac{1 - \hat{P}_{it}^{IIC|C}}{\hat{P}_{it}^{IIC|C}} \hat{h}_{it}^{OC}, \tag{9}$$

where we use the estimate $\hat{h}_{it}^{C}$ from Eq. (6). We estimate $\hat{P}_{it}^{IIC|C}$ from a probit model, in which we estimate the probability of IIC among SHARE community residents on a quadratic polynomial in the IADL-limitation count, country dummies, and interactions of these. We obtain $\hat{h}_{it}^{OC}$ from a regression of OC hours on a quadratic polynomial of the IADL-limitation count, country dummies, and interactions of these. We opt for a quadratic polynomial instead of the full set of dummies for the IADL-limitation count since the number of observations is low at the country level in SHARE and some IADL-count/country cells would otherwise be empty.

Table 12 shows how the imputation of IIC hours plays out for the Dahl estimator. In all countries, the estimate for community hours, $\hat{h}^C$, lies above the observed hours for the OC sample. Thus, our procedure infers that the IIC hours must be larger, as was to be expected. The lower the percentage of IIC recipients in the community, the stronger is this effect, which is why the Nordic countries have the largest estimates for IIC hours.

This can also be observed in Table 13, in which we compare the IIC-hours imputations from the Dahl selection model to the two OLS schemes. With respect to the pooled-OLS imputations, the selection model imputes high IIC hours in countries with a low IIC probability (the Nordic countries and France). In Austria and Germany, the IIC probality is somewhat higher and the Dahl estimate $\hat{h}^C$ lies closer to observed OC hours, thus the correction is smaller. Finally, the table shows that the grouped OLS scheme imputes about one hour more per country than pooled OLS; in comparison to the Dahl selection model, grouped OLS usually assigns lower or similar hours.

TABLE 12: Dahl imputation of IIC hours

Country	Community hours Dahl ($\hat{h}^C$)	OC hours (observed)	% IIC in community (observed)	IIC hours Dahl ($\hat{h}^{IIC}$)
Netherlands	2.34	1.28	0.23	6.84
Sweden	3.40	1.81	0.19	8.58
Denmark	2.91	1.45	0.18	8.71
Belgium	3.66	1.75	0.30	8.10
France	4.21	2.39	0.30	8.45
Germany	4.24	3.07	0.38	5.88
Austria	4.02	2.97	0.41	5.79
Spain	6.45	3.68	0.67	7.85
Italy	5.62	4.25	0.57	6.84
US	4.70	3.52	0.58	–

Table shows averages of hours over countries. SHARE sample restricted to Waves 1-2. Samples: all community residents (col. 1 and 3), all community residents not receiving IIC (col. 2), all community residents receiving IIC (col. 4).

TABLE 13: Comparison of imputation schemes for IIC hours

	Imputed IIC hours			Correction	
Country	Dahl ($\hat{h}^{IIC}$)	Grouped OLS	Pooled OLS	Dahl vs. grouped OLS	**Dahl vs. pooled OLS**
Netherlands	6.84	5.48	4.53	1.36	**2.31**
Sweden	8.58	5.54	4.38	3.04	**4.20**
Denmark	8.71	6.36	5.16	2.35	**3.55**
Belgium	8.10	6.58	5.34	1.52	**2.76**
France	8.45	6.58	5.37	1.87	**3.08**
Germany	5.88	6.38	5.27	-0.50	**0.61**
Austria	5.79	6.37	5.10	-0.58	**0.69**
Spain	7.85	7.49	6.18	0.36	**1.67**
Italy	6.84	6.85	5.52	-0.01	**1.32**

Table shows averages of IIC hours imputations across countries in first three columns, last two columns are differences between col. 3 and col. 1 and 2, respectively. SHARE sample restricted to Waves 1-2.

I.8. Assigning total hours to different care categories

Once we have imputed total hours of care, we assign them to the different care categories (IIC, OIC, FHC) in case that this information is missing. For nursing-home residents, we assume that they did not receive any informal care. Let h_{tot} be total hours of care, which equals $\hat{h}_{it}^{IIC}$ or $\hat{h}_{it}^{OC}$ depending on whether the respondent has an inside-household caregiver or not. We proceed as follows:

1. If there is only one type of care reported, all hours are assigned to this type of care. In the case of nursing-home residents, we assign all hours to NHC and set numbers in the other care categories to zero.

2. If exactly two types of care are reported (e.g. IIC and FHC) and...

   (a) ...hours are known for none of them: We code hours in the two categories as missing.

   (b) ...hours are known for exactly one type of care. Say h_i hours are known for care category i. We assign $\min\{h_i, 0.9 \times h_{tot}\}$ to category i, and $\max\{h_{tot} - h_i, 0.1 \times h_{tot}\}$ to the other care category. This means that we assign at least 10% of care hours to the smaller category in case the subtraction yields a value below 10%, which avoids negative hours.

   (c) (...hours known for both: No need to impute)

3. If the respondent reports all three types or care and...

   (a) ...hours are known for none of them: Code hours in all three categories missing.

   (b) ...hours are known for exactly one category. Say h_i hours are known for care category i: We then assign $\min\{h_i, 0.9 \times h_{tot}\}$ to category i and assign missing values to hours in the other two categories.

   (c) ...hours are known for exactly two care categories, say h_i and h_j:

      i. If $h_i + h_j \leq 0.9 \times h_{tot}$, we assign h_i and h_j to categories i and j and $h_{tot} - h_i - h_j$ to the remaining category.

      ii. If $h_i + h_j > 0.9 \times h_{tot}$, we assign $\frac{h_i}{h_i+h_j} 0.9 \times h_{tot}$ to i, $\frac{h_j}{h_i+h_j} 0.9 \times h_{tot}$ to j, and $0.1 \times h_{tot}$ to the third category.

   (d) (...hours known for all categories: No need to impute)

I.9. Model selection: A horse race

Our approach to model selection is based on comparing the accuracy of various models in predicting observed hours out of sample. In order to do so we utilize U.S. data since the selection model has predictions on community and nursing home residents, but not separately for IIC and OC (OIC+FHC) hours; in the SHARE data, however, IIC hours are not observed. In order to estimate the selection model, we use region as the instrument in the selection equation. In the second stage we use hours of care always leaving out one region which we then subsequently use to perform the out-of-sample test. Table 14 shows the results of this exercise. We use the average absolute error to assess the accuracy of prediction. In terms of predicting mean hours out of sample, all three models perform well. When including all five regions, the selection model does best; when excluding the region "Other" in calculating the mean absolute error the grouped OLS performs best.

TABLE 14: Out-of-sample predicted hours for U.S.

U.S. region	**actual hours**	**Dahl selection**	**pooled OLS**	**grouped OLS**
Northeast	4.494	4.409	4.411	4.533
Midwest	4.025	4.552	4.391	4.568
South	5.031	5.183	4.688	4.884
West	4.743	4.623	4.589	4.793
Other	6.063	5.860	5.653	5.421
Mean abs. error		0.217	0.271	0.285
(leaving out Other)		0.221	0.237	0.195

The table compares actual hours of care received in the community with those predicted by three candidate models for imputations (run on the four remaining regions).

For our European countries we cannot perform the same out-of-sample test as for the U.S. since the only hours available here are for OC.

I.10. Additional checks for Assumption 3

The fact that the regression models do well in predicting actual hours of care in the U.S. by region lends credence to the empirical plausibility of Assumption 3 that care hours conditional on a profile of functional limitations have the same expected value across regions. In order to provide additional tests, we also check (a) how actual hours available in SHARE compare to actual hours in the HRS, and (b) how community hours predicted compare with actual OC hours in Europe. Table 15 compares actual (non-imputed) hours of care in SHARE with actual hours of care in the HRS. As we note in the main text, we would not expect these numbers to be the same due to selection. In fact, we can see that for lower levels of dependency the hours are more similar but diverge for

larger numbers of functional limitation consistent with selection being more pronounced in Europe than the U.S. Table 16 shows that the Dahl selection model comes closer than the grouped OLS model to actual hours of care in Europe, while pooled OLS being farthest away from the data. The selection model yields predictions which are always above the actual hours (except in Austria), which is consistent with community residents being twice positively selected in Europe, once into the community (versus nursing home) and again within the community into OC (versus IIC).

TABLE 15: Actual hours in SHARE and HRS by number of limitations

dependency count	**SHARE**	**HRS:IIC+OC**	**HRS:OC**
1	1.48	1.63	1.08
	(585)	(1803)	(685)
2	2.44	2.21	1.67
	(307)	(1465)	(625)
3	2.83	3.09	2.03
	(200)	(1151)	(504)
4	3.65	3.81	2.51
	(128)	(938)	(430)
5	3.39	5.16	4.08
	(82)	(784)	(338)
6	5.43	6.32	4.14
	(59)	(626)	(275)
7	2.42	7.46	6.20
	(38)	(502)	(214)
8	5.57	8.84	7.12
	(29)	(391)	(176)
9	9.62	10.68	8.63
	(14)	(309)	(113)
10	8.17	11.89	10.51
	(14)	(244)	(74)
11	11.01	13.87	13.15
	(24)	(260)	(92)
12	14.77	16.04	15.56
	(6)	(173)	(61)
Total	2.91	4.65	3.28
	(1486)	(8646)	(3587)

The table compares actual hours of care available in SHARE with actual hours of care in the HRS by the number of functional limitations. In SHARE this information is available only for outside care (OC), which is outside-informal care and/or formal-home care, in waves 1 and 2. In the HRS, inside-informal care (IIC) is also available and so we provide two columns, one with all community residents and the other with only those respondents who obtain exclusively OC. We also report the number of individuals per number of functional limitation.

I.11. Data availability

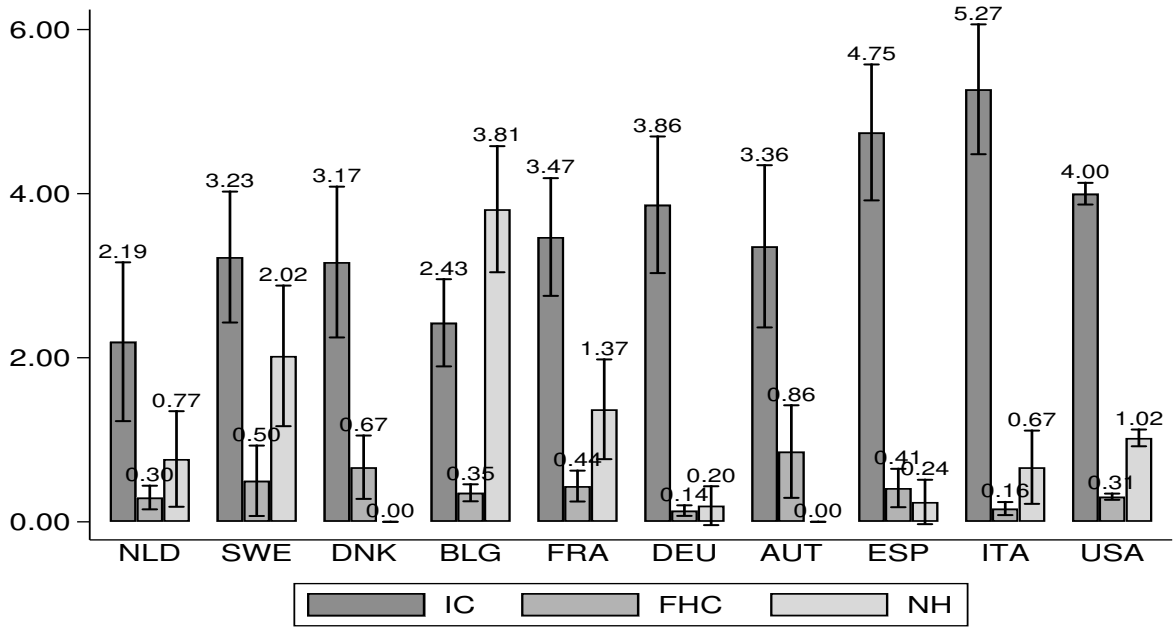
Codes for the above imputations and the resulting data tables are available on request from the authors. They will be made publicly available upon publication.

TABLE 16: Out-of-sample predicted hours for E.U.

country	actual hours	Dahl selection	pooled OLS	grouped OLS
Austria	2.973	2.819	2.663	2.132
Germany	3.067	3.184	2.979	2.423
Sweden	1.806	3.217	3.618	3.032
Netherlands	1.283	1.888	2.866	2.283
Spain	3.683	4.225	3.658	3.035
Italy	4.246	4.460	3.617	2.964
France	2.393	3.410	3.231	2.677
Denmark	1.452	2.381	3.012	2.422
Belgium	1.748	2.817	2.827	2.269
Mean abs. difference		0.673	0.880	0.824

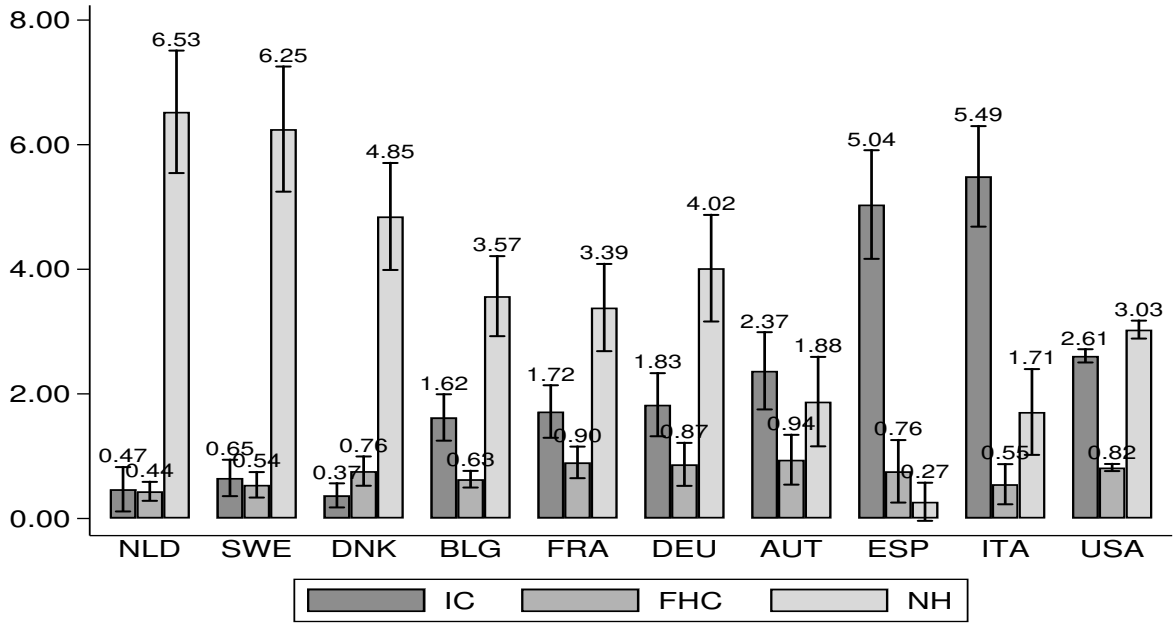
The table compares actual hours of outside-household care with those predicted by three candidate models for imputations.

FIGURE 12: Care choice: couples



Sample: All coupled respondents aged 65 and above who received care. SHARE: Waves 1-2. HRS: Waves 5-11. Weights are used.

FIGURE 13: Care choice: singles



Sample: All single respondents aged 65 and above who received care. SHARE: Waves 1-2. HRS: Waves 5-11. Weights are used.