



Published in final edited form as:

*Aging Ment Health*. 2017 March ; 21(3): 224–231. doi:10.1080/13607863.2015.1128881.

## Exploring the Benefits of Respite Services to Family Caregivers: Methodological Issues and Current Findings

Steven H. Zarit<sup>1</sup>, Yin Liu<sup>1</sup>, Lauren R. Bangert<sup>1</sup>, and Michael J. Rovine<sup>1</sup>

<sup>1</sup>Department of Human Development and Family Studies, The Pennsylvania State University

### Abstract

**Objectives**—There is growing emphasis on empirical validation of the efficacy of community-based services for older people and their families, but research on services such as respite care faces methodological challenges that have limited the growth of outcome studies. We identify problems associated with the usual research approaches for studying respite care, with the goal of stimulating use of novel and more appropriate research designs that can lead to improved studies of community-based services.

**Method**—Using the concept of research validity, we evaluate the methodological approaches in the current literature on respite services, including adult day services, in-home respite and overnight respite.

**Results**—Although randomized control trials (RCTs) are possible in community settings, validity is compromised by practical limitations of randomization and other problems. Quasi-experimental and interrupted time series designs offer comparable validity to RCTs and can be implemented effectively in community settings.

**Conclusion**—An emphasis on RCTs by funders and researchers is not supported by scientific evidence. Alternative designs can lead to development of a valid body of research on community services such as respite.

### Keywords

Respite; adult day care; quantitative methods; caregiving

---

Family care of persons at home can become an all-consuming enterprise that engulfs caregivers' daily life. From the earliest writings on family care, clinicians and caregivers themselves articulated the need for receiving breaks from the continual demands of providing care (e.g., Zarit, Orr, & Zarit, 1985). Respite represents a category of services that provide an alternative care arrangement so that the family caregiver can have time away from care responsibilities. Caregivers may use respite time to rest, attend to personal needs or in some cases continue their employment (Zarit, Stephens, Townsend, & Greene, 1998). Respite is offered in different ways, including in-home care, adult day services and overnight care. In contrast to time-limited psychoeducational interventions for caregivers, which tend to have short-lived benefits (Sørensen, Pinquart, & Duberstein, 2002), respite care can be sustained over time, helping to make home care a realistic alternative for families who prefer that option over institutionalized care.

Despite the potential advantages of respite care, research on its efficacy has been extremely limited and has not produced a substantial body of empirical evidence on outcomes, either for caregivers or for persons receiving respite. A recent comprehensive review of the literature on adult day services (ADS) from 2000 to 2011 found only two randomized control trials (RCTs) on outcomes for family caregivers and another 18 articles that described quasi-experimental studies that provided some type of control (e.g., matched sample) for evaluating the effects of the intervention (Fields, Anderson, & Dabelko-Schoeny, 2014). A Cochrane review (Maayan, Soares-Weiser, & Lee, 2014) that surveyed the literature as far back as 1989 identified four randomized trials of respite, although one of the four (Lawton, Brody, & Saperstein, 1989) was actually a trial of care management and not respite. We conducted our own search covering the period 1999 to the present (December, 2014) using the following data bases: PubMed, CAB Direct, and PsycINFO, and the search terms: *adult day, adult day care, adult day services, adult day health services, respite services, in-home care, and overnight respite*. The search identified 270 peer-reviewed articles on respite services. Focusing on studies that reported outcomes on family caregivers, we found 5 randomized control trials, 12 quasi-experimental studies and 4 within-person designs (findings available on request). Most studies had relatively small sample sizes ( $M = 84$ ,  $sd = 62$ , range = 16 – 212), except for one study with a particularly large sample size of 819 (Iecovich & Biderman, 2013). Indeed, the literature search identified more review papers ( $N = 32$ ) than outcome studies ( $N = 21$ ).

Given the promising benefits of respite to caregivers, we focus this paper on specific methodological challenges involved in respite research. Our goal is to suggest directions for more vigorous and innovative research on respite using methods suited for field studies. Some of the argument and proposals we make are not new, and indeed have their origins in now-classic works on methodology (e.g., Cook & Campbell, 1979), yet these approaches have not been widely adopted or accepted in the study of outcomes of respite and other community-based services.

Given the diversity in care situations and types of respite, we have intentionally limited the scope of the paper in several ways. First, we draw examples mainly from studies of one type of respite, adult day services (ADS). Second, as there have been surprisingly few studies of outcomes for persons receiving respite, we focus on family caregivers. We believe, however, that the approaches we propose have relevance for any type of respite, for studying outcomes of persons receiving respite and for research on the effectiveness of other types of community-based interventions.

## Research Challenges in the Study of Respite

The evaluation of respite care epitomizes translational research that seeks to systematically test practical questions in community settings. It has long been noted that the exigencies of community settings require adaptation and flexibility of methodological approaches, including design and measures, and that randomized control trials (RCT) may not be the optimal approach (e.g., Cook & Campbell, 1979; Jones, 1974; Shadish, Cook, & Campbell, 2001). The practical as well as ethical difficulties in implementing RCTs in community settings have been a barrier to development of a larger body of research on respite care,

while the assumption that only RCTs can generate valid results has blinded researchers and funders to alternative research strategies.

To examine the utility of RCTs and alternative designs in community studies, we organize our discussion around the concept of validities of research (Shadish, et al. 2001; Zarit, Stephens, & Femia, 2003). We consider challenges and limitations of RCTs and suggest alternative designs that can address validities well, so that confidence in the findings can approximate or even exceed an RCT. Where available, we provide examples of these alternative designs from the literature.

### The Validities of Research

Validity refers to the accuracy of the inferences from research findings (Shadish et al., 2001). Four main dimensions of validity are summarized in Table 1. RCTs are considered the “gold standard” of treatment designs based on the assumption that they provide more protection against threats to validity than alternative designs. The core belief about the superiority of RCTs is that *random* assignment produces *equivalent* treatment and control groups (i.e., internal validity) so that differences that emerge can be inferred as due to the effects of treatment, rather than to measured or unmeasured differences between groups. The equivalence of treatment and control groups in RCTs is *assumed*, most often not assessed and may not actually be achieved. Furthermore, when used in community-based trials, rather than in more tightly controlled laboratory or medical settings, RCTs encounter challenges that can compromise equivalence of treatment and control groups including small sample sizes, differential attrition and inadequate operationalization of the treatment.

**Sample size**—While not a consideration in large medication RCTs, sample size is often small in respite and other community studies. When sample size is small, differences between treatment and control groups are more likely to occur during recruitment even with randomized assignment. Differences between groups in characteristics such as gender, kin relationship (adult child versus spouse), socioeconomic status or employment status all have implications for the caregiver’s experience and could affect recruitment, retention and outcomes. For example, compared to daughters, caregivers who are wives who use ADS have provided care for a longer period before starting ADS use and show less improvement in subjective stress and positive emotion (Cho, Zarit, & Chiriboga, 2009; Kim, Zarit, Femia, & Savla, 2012). Likewise, characteristics of the person receiving care, such as type of disability or willingness to use services, can affect caregivers’ outcomes.

These types of *a priori* group differences can be statistically controlled through methods such as analysis of covariance (ANCOVA). Whether statistical control adequately addresses these differences, however, has been a controversial topic as long as ANCOVA has been used. More modern approaches to statistical control have emphasized the benefits of matching to address lack of equivalence, most elegantly presented through propensity score modeling (Pearl, 2009, Rosenbaum & Rubin, 1983).

**Differential attrition**—A more fundamental threat to the internal validity of RCTs in community settings is differential attrition following treatment assignment. Differential attrition, where characteristics of dropouts in the treatment and control groups differ can lead

to biased estimates of treatment effects (Heinsman & Shadish, 1996). The potential for differential attrition is considerable in RCTs of community programs. In contrast to RCTs involving medications, there often is no plausible placebo in a respite study. Typical practice is to randomly assign people into a treatment group where they receive the service immediately or to a control group where they are placed on a wait list or receive a minimal intervention (e.g., handing people a list of services available in the community). This has proven to be a difficult design to implement, with problems in initial recruitment of samples and differential attrition (e.g., Buckwalter et al., 1999). Participants not receiving services are prone to drop out or for crossover effects, where they use the same or equivalent services in the community (Lawton et al., 1989). Furthermore, a wait list control is feasible and ethical only for a short period of time, yet the benefits of respite and other community programs are likely to emerge only over longer periods of time (Zarit et al., 1998). The alternative, maintaining people in a control condition for a long period of time, is likely to result in considerable attrition as well as what Cook and Campbell (1979) call “resentful demoralization.”

**Inadequate operationalization and implementation of treatment**—Another consideration that has not received adequate attention in the respite literature is construct validity pertaining to the adequacy of operationalization and implementation of the treatment. Specifically, in RTCs of community services, treatment groups have often been offered low levels of services that may not be sufficient to reduce caregiver burden or distress (Baumgarten, Lebel, Laprise, Leclerc, & Quinn, 2002; Newcomer, Fox, & Harrington, 2001). Further contributing to low service use, researchers have sometimes recruited people based on their role incumbency as caregivers, and not because they were seeking respite (Lawton et al., 1989). Then, after caregivers are enrolled in the study, some are offered respite services. Not surprisingly, caregivers who were not seeking respite in the first place use sub-therapeutic amounts of service.

When there are no insurmountable practical or ethical barriers, RCTs offer an efficient and well-accepted model, but given that RCTs do not work well under some circumstances, we need to consider other designs that offer comparable validity.

### Other Threats to Validity in Respite Research

Before turning to alternate designs, we briefly mention other threats to validity that are pertinent to respite studies, whether in an RCT or another design.

**Restriction of range in outcome measures**—The assumption in treatment research has often been that all caregivers experience stress and experience negative emotions. When caregivers are recruited based on their role incumbency, the result is that a sizable proportion have low levels of the problems (e.g., depression, burden) that the treatment is designed to address (Belle et al., 2006; Quayhagen et al., 2000; Zarit & Femia, 2008). As a result, these caregivers cannot demonstrate improvement on the outcome measures and so the power to detect a treatment effect is reduced. A comparable problem occurs when an initial sample of caregivers is selected with high depression or subjective burden scores. In that case, there will likely be regression toward the mean over the course of the study in both treatment and

control groups, which may obscure differences due to effects of the treatment, particularly in small sample studies. These points underscore the heterogeneity among family caregivers and the need for more refined approaches to recruitment in treatment studies.

**Construct validity of the outcome measures**—Construct validity is usually discussed in terms of psychometric properties, but we can also ask if the right outcomes have been selected. Researchers need to give more thought about what outcomes we might reasonably expect of respite and as well as taking into account the goals of caregivers (Dabelko & Zimmerman, 2008; Zarit et al., 1998). Some specific outcomes of respite for caregivers include having time away from the care receiver, and reducing feelings of overload and strain in providing care.

**Use of a single method of assessment of outcomes**—Another threat to construct validity is the use of a single method of assessment. In respite and other caregiver studies, there is often reliance solely on self-report measures, which are prone to social desirability effects and response bias. This issue can be addressed by using multiple methods of assessment (e.g., biological markers of stress, observational measures of the caregiver and/or respite user, time use measures).

## Alternative Designs: Quasi-experimental Approaches

Quasi-experimental designs are an alternative when RCTs cannot be conducted without sacrificing validity or because of ethical considerations centered on withholding treatment for a control condition (Grimshaw, Campbell, Eccles, & Steen, 2000). There are many different types of quasi-experimental designs, but we focus on non-equivalent control group designs (Shadish et al., 2001). In these designs, one group receives the treatment and another group does not with no random assignment to groups (Table 2). Depending on the variant of non-equivalent control group design selected, assessments are conducted prior to and following treatment and may also include longer-term follow-ups. This type of design is regarded as inferior to RCTs because of an assumed lack of equivalence of the treatment and control groups prior to treatment. But just as RCTs often lack equivalence, quasi-experimental designs can implement procedures that create a valid test of a program.

One approach is to select caregivers who have elected to use a service or are already using the service and then compared to individuals not using the service (i.e., Femia, Zarit, Stephens, & Greene, 2007; Jardim & Pakenham, 2009; Wilz & Fink-Heitz, 2008). This strategy has the advantage of reducing attrition likely to occur with random assignment of people who do not seek the service in the first place and also reducing the crossover effect of control participants who use comparable services. But self-selection introduces other potential sources of variance. For example, Wilz and Fink-Heitz (2008) described the distribution of participants into a treatment or control group as a “first-come, first-served” process which in turn leads to unequal control and treatment groups. Matching the control group with the treatment group on characteristics associated with service use, particularly pre-treatment levels on outcome measures, is a standard way of addressing this problem. Matching is readily employed due to the ease at which researchers can collect this type of objective data (i.e., Zank, & Schacke, 2002). For example, Dröes, Breebaart, Meiland, van

Tilburg, and Mellenbergh (2004) conducted active matching on severity of dementia and characteristics that were associated with caregiver burden such as behavior problems of the person with dementia, the level of assistance required, and the competence of the caregiver. Other variables that might be used for matching or covarying differences include demographic characteristics such as education and income, living arrangement of participants, gender, and type of relationship between caregiver and care receiver (e.g., spouses, parent/child). Propensity score modeling provides a mechanism for strengthening inferences when applied to designs such as non-equivalent control group designs in which random assignment to treatment and control does not occur (Pearl, 2009; Rosenbaum & Rubin, 1983).

There remains controversy about how well these procedures work to reduce threats to validity. At the end of their Cochrane review on respite, Maayan and colleagues (2014) present the traditional view that even when treatment and control groups are roughly equivalent due to sampling methods and/or statistical approaches, quasi-experimental designs cannot approach RCTs in their validity. They state that RCTs have the “advantage of randomization that as well as controlling for factors that are known to affect relevant outcomes it controls for factors that are not known” (p. 19). On the other hand, Rosenbaum (2010) argues that with post-hoc matching procedures such as propensity score modeling, relatively unbiased estimates of treatment/control differences are not only possible, but approach the validity of an RCT.

Heinsman and Shadish (1996; Shadish, 1997) have provided compelling empirical evidence regarding the equivalence of RCT and quasi-experimental designs. They conducted a meta-analysis of 51 randomized and 47 non-randomized trials covering selected topics in educational, health education and drug abuse prevention trials. The results indicated that RCTs yielded larger effect sizes of treatment than quasi-experimental designs. These differences in effect sizes were largely due to other features of the design than how treatment was assigned. Specifically, studies using a passive (i.e., no treatment) control group had a larger effect size, and these passive control groups were found more commonly in RCTs. Other factors that accounted for the differences were self-selection into the treatment groups, attrition rates and pre-treatment differences in the outcome variables. Controlling for these variables reduced the difference in influence of random assignment on treatment effect size to a small amount (95% confidence interval between 0.176 and -0.016; Heinsman & Shadish, 1996). Shadish has replicated these results in other contexts (Shadish, Clark & Steiner, 2008; Shadish, 2011; Shadish, Galindo, Steiner, Wong, & Cook, 2011). Other researchers have similarly shown the relative efficacy of the quasi-experimental approach as compared to the randomized experiment (Benson & Hartz, 2000; Raaijmakers, Koffiberg, Posthumus, van Hout, van Engeland, & Matthys, 2008; Steiner, Cook, Shadish, & Clark, 2010).

These findings have two main implications. First, as Heinsman and Shadish (1996) recommend, researchers who address potential confounding factors in a quasi-experimental design will obtain findings that closely approximate an RCT. Recommended steps include controlling attrition, not allowing self-selection or minimizing its impact by selecting a control group from a similar population, and either minimizing pre-test differences on

outcome variables or adjusting post-test scores for any pre-test differences. It would also be important to control for effects of attention or expectations in the control group, a consideration that applies equally to RCTs and quasi-experimental designs. The second implication is that a well-designed quasi-experimental design will produce more accurate results than a poorly designed or compromised RCT. Furthermore, any bias due to the use of quasi-experimental designs would be in the direction of failing to detect a treatment that is effective, rather than reporting false positive findings (Rosenbaum & Rubin, 1983). In addition, provided that bias is due to confounding by indication, which is defined as an individual's probability of receiving the treatment with the intervention of interest given the complete set of all information known about that individual, much of the bias can be removed statistically and the effect size of the quasi-experiment comes close to that of the RCT (Rosenbaum & Rubin, 1984). Thus, there is little basis to the concern that quasi-experimental methods will affirm all sorts of questionable treatments that would not be found effective if an RCT were performed.

### **A Quasi-experimental Design of Effects of Adult Day Care Use on Family Caregivers**

We present an example of a non-equivalent control group quasi-experimental design that was used in one of the larger studies to date of the benefits of ADS for family caregivers of persons with dementia (Zarit et al., 1998). The treatment group consisted of caregivers who were enrolling a relative with dementia into an adult day care program. Initial assessments took place prior to enrollment. The control group was drawn from another geographic region in the USA that had similar social characteristics (e.g., population density, education, income) but a very limited network of adult day care programs for persons with dementia. As Heinsman and Shadish (1996) note, recruiting controls from another locale with a similar population reduces the effects of self-selection into treatment. To control further for factors that might be associated with the propensity to use services, prospective participants in the control communities were screened for their interest in receiving respite. Only those people in the control group who said they would consider using adult day care for their relative if it were available and affordable were included in the study.

Other procedures were used to assure equivalence of the samples. All caregivers had to be the primary person providing care to their relative with dementia, as indicated by the amount of care they provided and having responsibility for making decisions about care. Since the goal of the study was to determine the effects of receiving respite, people in both the treatment and control group who were already receiving more than a minimal amount of in-home paid respite (i.e., > 8 hours) were not enrolled in the study. For the treatment group, this meant that the sample would not include people who were simply substituting adult day care for another type of paid respite and thus would be unlikely to show any change with ADS use. It also eliminated people from the control group who were, in effect, receiving comparable amounts of a respite intervention as the treatment group.

The study also addressed construct validity of the outcome measures. In addition to the usual focus on affect, measures were selected that addressed the anticipated effects of adult day care, specifically, that the time away that caregivers had from the person with dementia would reduce time pressures, effort and energy expended in relation to caregiving. These

processes were assessed with measures of role overload, role captivity and caregiver strain (Pearlin, Mullan, Semple, & Skaff, 1990). Finally, pre-treatment scores and demographic variables were used as covariates in analyses.

The findings showed that, compared to the control group, caregivers in the treatment group had reduced overload, strain, depression and anger after three months of adult day care use, and reduced feelings of overload and depression after one year of use (Zarit et al., 1998). Additionally, adult day care use delayed institutional placement for daughters who were caring for a parent, but not for wives caring for a husband (Cho et al., 2009).

Another use of the non-equivalent control group design is to compare enhanced or innovative programs to standard care. As an example, Gitlin and colleagues (Gitlin, Reever, Dennis, Mathieu, & Hauck, 2006), compared an ADS program that integrated counseling and care management for family caregivers with standard programs, and found that the enhanced program resulted in improved well-being and lower nursing home use. As programs for caregivers become more common, this type of comparison can be used to identify the most promising approaches.

### **The Removed Treatment and Reversal Quasi-experimental Designs**

In contrast to models that examine group differences, several treatment designs use persons as their own controls. The traditional issue of “equivalence” is moot, since people are compared to themselves. Furthermore, procedures are used to demonstrate that outcomes are directly influenced by treatment and are not due to placebo effects. We look first at the removed treatment and reversal designs, which are generally categorized as “quasi-experimental” approaches (Barlow et al., 2009; Shadish et al., 2001), and then consider two extensions of these models, Interrupted Times Series and Person-Specific Models.

As shown in Table 2, removed treatment and reversal models surround treatment ( $X_+$ ) with pre- and post-observations ( $O_1$  and  $O_2$ ). Varying these sequences can lead to different types of comparisons of outcomes when people are exposed to treatment and when treatment is not present. For example, adding a follow-up post observation ( $O_3$ ) allows a second “pre/post” set of observations with no intervening treatment. The pre-post with intervening treatment can then be compared with the pre-post with no intervening treatment. Additional administrations of the treatment, sometimes called a repeated treatment design can show whether the effect of the treatment builds, remains constant or trails off. The removal of treatment after its initial presentation can also be considered a reversal treatment (a treatment that serves to counteract the treatment). Changes that occur in response to the treatment, but trail off when it is removed can be assumed to be under the influence of the treatment.

The way in which respite is usually administered follows the pattern of the treatment reversal design. Caregivers have days during which they receive respite and when they do not receive it and instead provide most or all of the care. Within-person effects of treatment can be evaluated by comparing each person’s outcomes on respite and non-respite days. An advantage of this approach is that it does not confound within-person changes with between person differences in levels on an outcome measure. In other words, one caregiver might



have a low score for depressive symptoms on a respite day and another might have a higher score, but both of them could have improved compared to their respective scores on non-respite days.

### **A Treatment Removed Design for Caregivers: The Daily Stress and Health (DaSH) Study**

The Daily Stress and Health (DaSH) Study is an example of a treatment removed design (Zarit, Kim, Femia, Almeida, & Klein, 2014). The study also has the advantage of using multiple methods for assessing outcomes: caregiver reports and biological markers of caregivers' stress response. The sample consisted of caregivers of persons with dementia who were using ADS. Caregivers were interviewed on 8 consecutive days. The idea was to be able to compare caregivers to themselves on days they used ADS for their family member and days they did most or all of the care. Building on prior work that found a large difference in caregivers' stressor exposure on ADS compared to non-ADS days (Zarit, Kim, Femia, Almeida, Savla & Molenaar, 2011), it was expected that caregivers would demonstrate improved subjective stress, affect and biomarkers of the stress response. For each of the 8 observation days, caregivers were interviewed by telephone in the evening and answered questions about care and non-care stressors that occurred in the past 24 hours and about their affect and health symptoms. They also provided 5 saliva samples each day, in order to assess biological indicators of the stress process. Despite the intensive nature of data collection, attrition was low. Out of 200 caregivers who enrolled in the study, 173 (86.5%) completed most or all of the daily interviews and saliva samples (Zarit et al., 2014).

Another notable feature is the design of the study lends itself to assessing treatment effects at the point when they are likely to be strongest—during or immediately following use of respite. Nearly all studies of respite have taken a conventional approach to assessment of outcomes; that is, asking caregivers to report feelings over some previous period, usually one week, but sometimes longer, that includes *both* respite days and non-respite days. In effect, caregivers would be averaging their experience across high-stress (caregiving) days and low-stress (respite) days (Zarit et al., 2011). This process can be affected disproportionately by the most recent events, or by one particularly stressful event that occurred during the recall period. Daily or more frequent assessments reduce recall bias because they occur in proximity to when caregivers use or do not use respite.

Turning to the findings, ADS use reduced caregivers' feelings of anger and buffered the effects of both daily care-related stressors and non-care stressors on depressive symptoms (Zarit et al., 2014). Furthermore, more days of ADS use of the period of observation was associated with less daily emotional reactivity overall (Liu, Kim, Almeida, & Zarit, 2015). In addition to subjective reports, caregivers showed better regulation of two stress hormones, cortisol and dehydroepiandrosterone-sulfate (DHEA-S) in response to ADS days (Klein, Kim, Almeida, Femia, Rovine & Zarit, 2014; Zarit, Whetzel, Kim, Femia, Almeida, Rovine, & Klein, 2014). For cortisol, effects were found on the same day as ADS use, while DHEA-S showed positive responses (increased levels) on days following ADS use. Cumulative effects of ADS use were found for health outcomes one year later. Caregivers who received more days of respite were less likely to experience declines in functional health compared to caregivers using fewer days of adult day care (Liu, Kim, & Zarit, 2015).

One concern about applications of this design beyond respite is the ethics of alternating administration and withdrawal of treatment. This should be done in ways that follows the natural course of administration of the intervention. It should also not lead to hardship among caregivers or undermine the initial effects of treatment.

## Within-Person Designs as a Paradigm Shift in Intervention Research

The DaSH study represents a growing interest in within-person comparisons as a research strategy. Perhaps the main advantage is that persons serve as their own controls. Trends such as “personalized medicine” are built on recognition that people vary in their responses to a treatment, whether medications or behavioral procedures, and consequently more individualized approaches to type of treatment and dosage are needed. Furthermore, methodologists have provided a growing rationale for more individual focused research strategies. Using mathematic models based on ergodic theorems (Birkhoff, 1931), Molenaar (2004; Molenaar & Campbell, 2009) demonstrates that individual processes of change cannot be inferred accurately from between-person comparisons of differences. Additionally, the larger and more representative the sample is of the total population, the less likely the findings will apply to any given individual (Barlow et al., 2009; Molenaar & Campbell, 2009).

Rather than studying group differences, within-person approaches can examine intraindividual changes across multiple people, using statistical methods appropriate to this type of data. This strategy can generate refined models of change processes that more accurately reflect the range and sources of individual variation in change (Molenaar, 2008; Velicer, Babbin, & Palumbo, 2014). Within-person approaches can also examine which treatment or combinations of treatment can be most effective for a given individual. Given the heterogeneity among caregivers and caregiving situations, use of these approaches could lead to more individualized tailoring of treatment as well as ability to adjust type or dosage of interventions depending on the caregiver’s responses and changing circumstances.

We want to mention two such approaches, the Interrupted Time Series Designs (ITSD) and Person-specific Models. ITSD differs from the treatment removed and reversal approaches in using multiple observations prior to, during and following treatment to establish levels and patterns of change of outcome measures affected by the treatment (see Table 2; Barlow et al., 2009; Shadish et al., 2001). As with the prior models, the inference that the treatment leads to changes in outcome measures is established by comparing the person’s scores when the person is not receiving treatment to scores when treatment has been administered. Multiple replications of re-introducing and withdrawing treatment (Table 2, R-ITSD Model 3) can extend and strengthen the evidence for the associations between treatment and outcomes. Each replication includes additional “posttests.” The expectation is for a pre-post difference (first presentation of the treatment), a smaller pre-post difference (removal of treatment), an increased pre-post difference (reintroduction of the treatment), and so on. If outcomes are affected in this way, it can be inferred that they are at least partly under the control of the treatment.

In contrast to ITSDs, which examine differences across individuals using the same statistical model to describe all participants, the person-specific approach (Molenaar, 2004; Molenaar & Campbell, 2009) creates individual models of change. The large number of occasions of measurement for each person makes it possible to estimate a separate time series model for each individual and test the assumption regarding the homogeneity of the models, that is, whether a common model used to assess the pre- and post-treatment evaluation is a good model for each participant in the study.

When qualitatively different models describe different individuals or subgroups of individuals, the results of pooling across groups can lead to a false representation of how the individual responds to the treatment (Nesselroade & Molenaar, 1999). One way to think of the effect of inappropriate pooling would be if one subgroup were strongly affected by the treatment and another had no effect. Depending on the relative sizes of the subgroups, the effect could disappear with pooling. The person-specific approach can be useful in developing refined models of change processes that more accurately reflect the range and sources of individual variation (Molenaar, 2008; Molenaar, Huizenga, & Nesselroade, 2003; Velicer et al., 2014).

The relatively large number of observations needed per person in both ITSD and person-specific approaches where essentially the number of observations would equal the number of individuals in a cross-sectional study limits practical application. However, both approaches provide a framework for developing a science of change that could inform models of treatment for caregivers and for other problems in mental health of later life.

## Conclusions

The emphasis on RCTs by funders and researchers as the sole valid treatment design is not supported by scientific evidence. Reliance on RCTs restricts the range of research that can be conducted in community-based settings, and also limits testing innovative models of change that could lead to more individualized treatment approaches. Non-comparable control group designs represent a useful tool that, when conducted properly, yield valid data about the efficacy of treatments, while avoiding some of the common threats to validity of RCTs conducted on respite and other community-based programs. Various types of within-person designs also offer promise, particularly for treatments such as respite that are administered intermittently. Within-person comparisons when caregiver or the respite client is using and not using respite can establish effects that treatment may have on outcomes. Additionally, multiple types of measures in respite studies (self-reports, observations, biological measures) allow for a more comprehensive examination of potential outcomes. Use of alternative designs to the RCT should not be an excuse for careless research or poor control strategies for evaluating outcomes. Rather, well-controlled studies that use alternative designs can expand the empirical foundation for respite as well as for many other community interventions.

## Acknowledgments

The research in this paper was supported in part by a grant R01 AG031758 “*Daily Stress and Health Study*” (PI: Steven H. Zarit, PhD) from the National Institute of Aging.

## References

- Barlow, DH., Nock, MK., Hersen, M. Single case experimental designs: Strategies for studying behavior change. 3. Boston, MA: Pearson/Allyn and Bacon; 2009.
- Baumgarten M, Lebel P, Laprise H, Leclerc C, Quinn C. Adult day care for the frail elderly: Outcomes, satisfaction, and cost. *Journal of Aging and Health*. 2002; 14:237–259. DOI: 10.1177/089826430201400204 [PubMed: 11995742]
- Belle SH, Burgio L, Burns R, Coons D, Czaja SJ, Gallagher-Thompson D, et al. Enhancing the quality of life of dementia caregivers from different ethnic or racial groups. *Annals of Internal Medicine*. 2006; 145:727–738. DOI: 10.7326/0003-4819-145-10-200611210-00005 [PubMed: 17116917]
- Benson K, Hartz K. A comparison of observational studies and randomized controlled trials. *New England Journal of Medicine*. 2000; 342:1878–1886. [PubMed: 10861324]
- Birkhoff G. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences, USA*. 1931; 17:656–660. DOI: 10.1073/pnas.17.2.656
- Buckwalter KC, Gerdner L, Kohout F, Hall GR, Kelly A, Richards B, Sime M. A nursing intervention to decrease depression in family caregivers of persons with dementia. *Archives of Psychiatric Nursing*. 1999; 13:80–88. DOI: 10.1016/S0883-9417(99)80024-7 [PubMed: 10222636]
- Cho S, Zarit SH, Chiriboga DA. Wives and daughters: The differential role of day care use in the nursing home placement of cognitively impaired family members. *The Gerontologist*. 2009; 49:57–67. DOI: 10.1093/geront/gnp010 [PubMed: 19363004]
- Cook, TD., Campbell, DT. *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin; 1979.
- Dabelko HI, Zimmerman JA. Outcomes of adult day services for participants: A conceptual model. *Journal of Applied Gerontology*. 2008; 27:78–92. DOI: 10.1177/0733464807307338
- Dröes RM, Breebaart E, Meiland FJM, van Tilburg W, Mellenbergh GJ. Effect of meeting centres support program on feelings of competence of family carers and delay of institutionalization of people with dementia. *Aging & Mental Health*. 2004; 8:201–211. DOI: 10.1080/13607860410001669732 [PubMed: 15203401]
- Femia EE, Zarit SH, Stephens MAP, Greene R. Impact of adult day services on behavioral and psychological symptoms of dementia. *The Gerontologist*. 2007; 47:775–788. DOI: 10.1093/geront/47.6.775 [PubMed: 18192631]
- Fields NL, Anderson KA, Dabelko-Schoeny H. The effectiveness of adult day services for older adults: A review of the literature from 2000 to 2011. *Journal of Applied Gerontology*. 2014; 33:130–163. DOI: 10.1177/0733464812443308 [PubMed: 24652952]
- Gitlin LN, Reeve K, Dennis MP, Mathieu E, Hauck WW. Enhancing quality of life of families who use Adult Day Services: Short- and long-term effects of the Adult Day Services Plus program. *The Gerontologist*. 2006; 46:630–639. DOI: 10.1093/geront/46.5.630 [PubMed: 17050754]
- Grimshaw J, Campbell M, Eccles M, Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Family Practice*. 2000; 17(Suppl 1):S11–S16. [PubMed: 10735262]
- Heinsman DT, Shadish WR. Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods*. 1996; 1:154–169. DOI: 10.1037/1082-989X.1.2.154
- Iecovich E, Biderman A. Attendance in adult day care centers of cognitively intact older persons: Reasons for use and nonuse. *Journal of Applied Gerontology*. 2013; 32:561–581. DOI: 10.1177/0733464811432141 [PubMed: 25474762]
- Jardim C, Pakenham KI. Pilot investigation of the effectiveness of respite care for carers of an adult with mental illness. *Clinical Psychologist*. 2009; 13:87–93. DOI: 10.1080/13284200903353064
- Jones, RR. Design and analysis problems in program evaluation. In: Davidson, PO, Clark, FW., Hamerlynck, LA., editors. *Evaluation of behavioral programs in community, residential, and school settings: Proceedings of the fifth Banff international conference on behavior modification*; Champaign, IL: The Research Press; 1974. p. 1-32.

- Kim K, Zarit SH, Femia EE, Savla J. Kin relationship of caregivers and people with dementia: Stress and response to intervention. *International Journal of Geriatric Psychiatry*. 2012; 27:59–66. [PubMed: 21322030]
- Klein LC, Kim K, Almeida DM, Femia EE, Rovine MJ, Zarit SH. Anticipating an easier day: Effects of adult day services on daily cortisol and stress. *Gerontologist*. 2014, epub.
- Lawton MP, Brody EM, Saperstein AR. A controlled study of respite service for caregivers of Alzheimer's patients. *The Gerontologist*. 1989; 29:8–16. DOI: 10.1093/geront/29.1.8 [PubMed: 2753374]
- Liu Y, Kim K, Almeida DM, Zarit SH. Daily fluctuation in negative affect for family caregivers of individuals with dementia. *Health Psychology*. 2015; 34:729–740. DOI: 10.1037/hea0000175 [PubMed: 25365414]
- Liu Y, Kim K, Zarit SH. Health trajectories of family caregivers: Associations with care transitions and adult day service use. *Journal of Aging and Health*. 2015; 27:686–710. DOI: 10.1177/0898264314555319 [PubMed: 25348275]
- Maayan, N., Soares-Weiser, K., Lee, H. Cochrane database of systematic reviews. Vol. 2. The Cochrane Library; 2014. Respite care for people with dementia and their carers.
- Molenaar PCM. A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*. 2004; 2:201–218. DOI: 10.1207/s15366359mea0204\_1
- Molenaar, PCM. Consequences of the ergodic theorems for classical test theory, factor analysis, and the analysis of developmental processes. In: Hofer, SM., Alwin, DF., editors. *Handbook of cognitive aging*. Thousand Oaks, CA: Sage; 2008. p. 90-104.
- Molenaar PCM, Campbell CG. The new person-specific paradigm in psychology. *Current Directions in Psychological Science*. 2009; 18:112–117. DOI: 10.1111/j.1467-8721.2009.01619.x
- Molenaar, PCM., Huizenga, H., Nesselroade, JR. The Relationship Between the Structure of Interindividual and Intraindividual Variability: A Theoretical and Empirical Vindication of Developmental Systems Theory. In: Staudinger, U., Lindenberger, U., editors. *Understanding human development*. New York, NY: Springer; 2003. p. 339-360.
- Nesselroade, JR., Molenaar, PCM. Pooling lagged covariance structures based on short, multivariate time-series for dynamic factor analysis. In: Hoyle, RH., editor. *Statistical strategies for small sample research*. Newbury Park, CA: Sage; 1999. p. 223-251.
- Newcomer RJ, Fox PJ, Harrington CA. Health and long-term care for people with Alzheimer's disease and related dementias: Policy research issues. *Aging and Mental Health*. 2001; 5(Suppl 1):S124–S137. DOI: 10.1080/713649997 [PubMed: 11513488]
- Pearl, J. *Causality*. 2. New York: Cambridge; 2009.
- Pearl, J. *Causality: Models, reasoning, and inference*. 2. New York, NY: Cambridge University Press; 2009.
- Pearlin LI, Mullan JT, Semple SJ, Skaff MM. Caregiving and the stress process: An overview of concepts and their measures. *The Gerontologist*. 1990; 30:583–594. DOI: 10.1093/geront/30.5.583 [PubMed: 2276631]
- Quayhagen MP, Quayhagen M, Corbeil RR, Hendrix RC, Jackson JE, Snyder L, et al. Coping with dementia: Evaluation of four nonpharmacologic interventions. *International Psychogeriatrics*. 2000; 12:249–265. DOI: 10.1017/S1041610200006360 [PubMed: 10937544]
- Raaijmakers M, Koffiberg H, Posthumus J, van Hout J, van Engeland H, Matthys W. Assessing performance of a randomized versus non-randomized study design. *Contemporary Clinical Trials*. 2008; 29(2):293–303. [PubMed: 17707138]
- Rosenbaum, PR. *Design of observational studies*. New York, NY: Springer; 2010.
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55. DOI: 10.2307/2335942
- Rosenbaum PR, Rubin DB. Reducing bias in observational in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*. 1984; 79:516–524. DOI: 10.1080/01621459.1984.10478078
- Shadish, WR. NIDA research monograph (1046–9516). National Institute on Drug Abuse; 1997. Experiments versus quasi-experiments: do they yield the same answer; p. 147

- Shadish WR. Randomized controlled studies and alternative designs in outcome studies: challenges and opportunities. *Research on Social Work Practice*. 2011; 21(6):636–643.
- Shadish WR, Clark MH, Steiner PM. Can Nonrandomized Experiments Yield Accurate Answers? A Randomized Experiment Comparing Random to Nonrandom Assignment. *Journal of the American Statistical Association*. 2008; 103:1334–1343.
- Shadish, WR., Cook, TD., Campbell, DT. *Experimental and quasi-experimental designs for generalized causal inference*. 2. Boston, MA: Wadsworth Cengage Learning; 2001.
- Shadish WR, Galindo RG, Steiner PM, Wong VC, Cook TD. A randomized experiment comparing random to cutoff-based assignment. *Psychological Methods*. 2011; 16:179–191. [PubMed: 21534684]
- Sörensen S, Pinquart M, Duberstein P. How effective are interventions with caregivers? An updated meta-analysis. *The Gerontologist*. 2002; 42:356–372. DOI: 10.1093/geront/42.3.356 [PubMed: 12040138]
- Steiner PM, Cook TD, Shadish WR, Clark MH. The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*. 2010; 15:250–267. [PubMed: 20822251]
- Velicer, WF., Babbin, SF., Palumbo, R. Idiographic applications: Issues of ergodicity and generalizability. In: Molenaar, P.Lerner, R., Newell, K., editors. *Handbook of relational developmental systems theory and methodology*. New York, NY: Guilford Publications; 2014. p. 425-441.
- Wilz G, Fink-Heitz M. Assisted vacations for men with dementia and their caregiving spouses: Evaluation of health-related effects. *Gerontologist*. 2008; 48:115–120. DOI: 10.1093/geront/48.1.115 [PubMed: 18381838]
- Zank S, Schacke C. Evaluation of geriatric day care units: Effects on patients and caregivers. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2002; 57(4):P348–P357. DOI: 10.1093/geronb/57.4.P348
- Zarit SH, Femia EE. A future for family care and dementia intervention research? Challenges and strategies. *Aging & Mental Health*. 2008; 12:5–13. DOI: 10.1080/13607860701616317 [PubMed: 18297475]
- Zarit SH, Kim K, Femia EE, Almeida DM, Savla J, Molenaar PCM. Effects of adult day care on daily stress of caregivers: A within person approach. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*. 2011; 66B:538–546. DOI: 10.1093/geronb/gbr030
- Zarit SH, Kim K, Femia EE, Almeida DM, Klein LC. The effects of adult day services on family caregivers' daily stress, affect and health: Outcomes from the DaSH Study. *The Gerontologist*. 2014; 54:570–579. DOI: 10.1093/geront/gnt045 [PubMed: 23690056]
- Zarit, SH., Orr, NK., Zarit, JM. *The hidden victims of Alzheimer's disease: Families under stress*. New York, NY: New York University Press; 1985.
- Zarit SH, Stephens MAP, Femia EE. The validity of research findings: The case of interventions with caregivers. *Alzheimer's Care Quarterly*. 2003; 4(3):216–228.
- Zarit SH, Stephens MAP, Townsend A, Greene R. Stress reduction for family caregivers: Effects of adult day care use. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 1998; 53B:S267–S277. DOI: 10.1093/geronb/53B.5.S267
- Zarit SH, Whetzel CA, Kim K, Femia EE, Almeida DM, Rovine MJ, Klein LC. Daily stressors and adult day service use by family caregivers: Effects on depressive symptoms, positive mood, and dehydroepiandrosterone-sulfate. *The American Journal of Geriatric Psychiatry*. 2014; 22:1592–1602. DOI: 10.1016/j.jagp.2014.01.013 [PubMed: 24566240]

**Table 1**

## Four Main Dimensions of Validity

<b>Validity Type</b>	<b>Definition</b>
Internal validity	The integrity of research procedures to produce accurate statements about the effects of a treatment or other manipulation, often stated as whether the empirical relationship between treatment and outcome can be considered causal
Construct validity	The accuracy of measures used to assess the domains or outcomes that are expected to change with treatment, and the accuracy of procedures represent the treatment construct
Statistical conclusion validity	The appropriateness of the statistical procedures used for determining the association between the treatment and outcomes.
External validity	Generalizability of the experiment to the population of interest and to other settings, treatments and outcomes

Source: Shadish et al. (2001).

**Table 2**  
Alternative Research Designs for Respite and Other Community-Based Treatments

Design	Advantages	Threats to Validity
<p><b>Non-equivalent Control Group Designs</b></p> <p>Group 1: <math>O_1 X_+ O_2</math>                      NR                      Group 2: <math>O_1 X_- O_2</math>                      One group receives the treatment (<math>X_+</math>) and another group does not (<math>X_-</math>) with no random assignment (NR) to group.</p>	<p>Comparison of treatment and control groups when random assignment is not possible.</p>	<p>Lack of equivalence of the groups. Can be controlled by reducing attrition, using an active control condition, covarying on potential differences that could affect outcomes, use of propensity score modeling to adjust for differences, selection of a control sample from a different setting or community where the treatment is not available.</p>
<p><b>Removed Treatment and Reversal Designs</b></p> <p><b>Single group removed treatment design</b>                      NR  <math>O_1 X_+ O_2 O_3 X_+ O_4</math>                      Treatment (<math>X_+</math>) is surrounded by pre- and post-treatment observations (<math>O</math>). The sequence of pre-post observation with intervening treatment can be repeated to assess if treatment effects are maintained over time.</p> <p><b>Single group reversed treatment design</b>                      NR  <math>O_1 X_+ O_2 O_3 X_- O_4</math>                      Treatment (<math>X_+</math>) is surrounded by pre- and post-treatment observations (<math>O_1 O_2</math>) and then removed (<math>X_-</math>) to determine if removal counteracts the effects of treatment</p>	<p>Can show if treatment effects are maintained over time when treatment is continued and/or removed.                      Effects that occur when treatment is administered and trail off when treatment is removed can be assumed to be influenced by the treatment.                      Participants can be compared to themselves when treatment is administered, continued and/or removed, avoiding the problem of lack of equivalence of control groups.</p>	<p>Selection of people seeking treatment can reduce generalizability to the broader population of interest.</p>
<p><b>Interrupted Time Series Designs</b></p> <p><b>Model 1. Basic Interrupted Time Series Design (ITSD)</b>                      Note: <math>O</math> = a set of one or more observations.  <math>O_1 O_2 X_+ O_3 O_4</math>  <b>Model 2. The Treatment Removal Interrupted Time Series Design (TR-ITSD)</b>  <math>O_1 O_2 X_+ O_3 O_4 X_- O_5 O_6</math>  <b>Model 3. Treatment Replication Interrupted Time Series Design (R-ITSD)</b>  <math>O_1 O_2 X_+ O_3 O_4 X_- O_5 O_6 X_+ O_7 O_8 X_- O_9 X_+ O_{10} O_{11}</math></p>	<p>Treatment effect to be compared against the no-treatment effect for the same person.                      If outcomes are affected during the treatment period, but revert closer to baseline during the withdrawal period, it can be inferred that the outcomes are at least partly under the control of the treatment</p>	<p>Small sample sizes can limit generalizability, though inferences about within person processes of change may be more reliable with this approach than in typical between person treatment-control designs.</p>

Notes: NR = non-randomized assignment;  $O$  = observation,  $X_+$  = exposure to treatment,  $X_-$  = no treatment or removal of treatment.