Children and Youth Services Review 104 (2019) 104390

Contents lists available at ScienceDirect

Children and Youth Services Review

journal homepage: www.elsevier.com/locate/childyouth

Psychological difficulties among custodial grandchildren

Gregory C. Smith[a,*], Bert Hayslip Jr[b], Britney A. Webster[a]

[a] College of Education Health and Human Service, Kent State University, Kent, OH, 44242, United States
[b] Department of Psychology, University of North Texas, Denton, TX 76203, United States

ARTICLE INFO

Keywords:
Custodial grandchildren
Psychological difficulties
Race
Gender

ABSTRACT

Although custodial grandchildren (CG) are likely to have more emotional and behavioral problems than children in general, only a handful of studies involving nationally representative data have investigated this important public health issue. The present study is unique in examining informant reports of psychological difficulties and prosocial behavior, obtained via the Strengths and Difficulties Questionnaire (SDQ) parent version, regarding two samples (n = 509 and n = 323) of CG between ages 4–12 and three samples of age peers from the 2004 National Health Interview Survey (NHIS) residing in homes with either no birth parent (n = 184), one parent (n = 1618), or both parents (n = 3752). A MANCOVA encompassing the main effects of sample type, child gender, and informant's race across six SDQ subscales (with informant age and education, as well as child age controlled) showed all three main effects to be statistically significant ($p < .001$). Between subjects effects for sample type were significant for all six subscales ($p < .001$), with a general pattern found whereby the two CG samples had higher difficulty and lower prosocial scores than all three NHIS samples, with corresponding effect sizes falling in the moderate to very large range. Tests of between subjects effects for child gender showed, that regardless of sample type, males typically had higher difficulty scores and lower prosocial scores than those for females. Tests of between subjects effects by informant's race showed only sporadic differences that were independent of sample type. Comparisons of the SDQ banded scores suggested that CG have considerably different cutoff points than do children in other family structures to indicate a likely diagnosis of a serious psychological disorder. We conclude that primary school aged CG are at an especially high risk for both internalizing and externalizing difficulties regardless of children's gender or informant's race.

1. Introduction

Nearly three million U.S. children are in homes where no parent is present, over half of whom live with a grandparent (Vandivere, Yrausquin, Allen, Malm, & McKindon, 2012). The United States Congress has reported that more than 2,500,000 grandparents in the U. S. are the primary caretaker of their grandchildren, with this number expected to increase as the opioid epidemic expands (GovTrack.Us, 2019. Children in these non-parental care arrangements represent an important public health concern given substantial evidence that living in a household with no biological parent present is often associated with poor physical and mental health outcomes (Bramlett & Blumberg, 2007; Vandivere et al., 2012; Ziol-Guest & Dunifon, 2014). Despite considerable research comparing the socio-emotional well-being of those children in out of home care (OOHC) who receive non-relative care versus kinship care (Washington et al., 2018), scant attention has been paid to the specific subgroup within kinship care known as custodial grandchildren (CG). The latter children are those being raised by grandparents in absence of birth parents irrespective of whether or not formal legal arrangements are made by grandparents such as adoption, foster care, or guardianship (Hayslip, Fruhauf, & Doolbin-MacNab, 2017).

We address this gap in the present study by examining the psychological difficulties in two different samples of CG between the ages of 4 and 12 as reported by their grandmothers who are raising them. We not only compare these samples to each other, but also to children of the same ages from the 2004 National Health Interview Survey (NHIS) who resided in households with either no biological parents, one biological parent, or both parents. These comparisons are important in view of prior findings that children are generally healthier in two-parent biological families, and that children in some non-marital family structures (e.g., single-mother and grandparent-only) have poorer physical and mental health (Bramlett & Blumberg, 2007; Vandivere et al., 2012; Ziol-Guest & Dunifon, 2014).

* Corresponding author.
E-mail address: gsmith2@kent.edu (G.C. Smith).

https://doi.org/10.1016/j.childyouth.2019.104390
Received 28 February 2019; Received in revised form 20 June 2019
Available online 25 June 2019

1.1. Why custodial grandchildren are susceptible to psychological difficulties

There are many reasons to suspect higher levels of psychological difficulties among CG. One is that they typically receive OOHC from grandparents due to such predicaments among their birth parents as substance abuse, child abuse and neglect, teenage pregnancy, death, illness, divorce, incarceration, and HIV-AIDS. In turn, these quandaries bear threats for psychopathology among CG, including exposure to prenatal toxins, early trauma, insufficient interaction with birth parents, family boundary ambiguity, uncertain futures, and social stigma (Smith & Palmieri, 2007). Many CG also experience frustration and anger toward their parents and grandparents, grief over the loss of parents, confusion over their grandparents' new role as parents, and sadness over moving away from friends (Edwards, 2009; Hayslip et al., 2017).

Like children receiving other types of OOHC kinship care, many CG reside in economically distressed homes and neighborhoods, with nearly one third living below the poverty line and being uninsured (Baker & Mutchler, 2010). Such economic hardship may maintain pre-existing, exacerbate current, or precipitate new CG psychological difficulties due to such forces as hostility and conflict in the home which adversely affect children's behavioral development, neighborhood impoverishment characterized by low quality schools and violence, and infrequent access to secondary caregivers to provide quality interactions with caring adults, adequate supervision, and needed stimulation (Edwards, 2009; Font, 2014).

Additional reasons for CG psychological difficulties are related to characteristics of the grandparents who provide care to them, the majority of whom are custodial grandmothers (CGM). There is evidence showing that CGMs experience high levels of psychological distress (believed to be associated with such stressors as grief, disappointment and anger toward CG birth parents, shame and social isolation, role overload and ambiguity, and difficulty in keeping up with the medical, financial, academic, and social demands of having to unexpectedly raise a growing child (Goulette, Evans, & King, 2016; Hayslip et al., 2017). In turn, the socioemotional distress felt by CGMs can adversely affect the CGs adjustment by impairing their parenting practices and creating a troubled emotional climate within the home (Conley, Caldwell, Flynn, Dupre, & Rudolph, 2004; Smith, Hayslip, Hancock, Strieder, & Montoro-Rodriguez, 2018). Even if not psychologically distressed, many CGMs are older than other caregivers and face socioeconomic disadvantages such as low educational attainment, insufficient income, racial discrimination, worsening health, and being unmarried (Hayslip et al., 2017; Vandivere et al., 2012). These disadvantages may adversely affect CG's behavioral development by restricting CGM's ability to provide proper levels of cognitive stimulation, affection, supervision, or basic resources (Font, 2014). Older CGMs are also likely to have outdated ideas about child development and discipline which may limit their parenting abilities (Hayslip et al., 2017).

1.2. Prior evidence of psychological difficulties in custodial grandchildren based on nationally representative survey data

Partial evidence for high levels of psychological difficulties among CG comes from studies with nationally representative survey data that compared children receiving non-parental kinship care of various types to either those in non-relative foster care or to those in the care of their biological parents. In general, children in non-parental kinship care fare worse on measures of mental and physical well-being than children in parental care, but better than those in non- relative care (Billing, Ehrle Macomber, & Kortenkamp, 2002; Rubin et al., 2008; Sakai, Lin, & Flores, 2011; Washington et al., 2018).

To date, only four studies involving nationwide data have explicitly compared the psychological well-being of CG to either children in other non-traditional care arrangements or to children within the general population. Smith and Palmieri (2007), for example, compared the emotional and behavioral difficulties of 733 CG between ages 4–17 to those of 9878 age peers sampled within the 2001 National Health Interview Survey (NHIS) as measured by the parent version of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997, 2001). They found that CG fared worse than peers from the NHIS sample across all domains measured by the SDQ regardless of the child's gender, and that more difficulties were reported for CG if CGMs were caring for boys, were White, or recruited via convenience versus population based methods.

Using data from the 2003 National Survey of Children's Health (NSCH), Bramlett and Blumberg (2007) examined the physical and mental health status of children under age 18 across six different family structures (two-biological parents, blended step, blended adoptive, biological mother only, biological father only, and grandparents only) while controlling for key sociodemographic differences. Children in grandparent only families not only differed significantly from children living with two biological parents, but they also had the poorest health status among any of the six family structures examined. Specifically, children in grandparent only families were almost twice as likely as those living with two biological parents to have special health needs or asthma-related health problems; more than twice as likely to have attention deficit disorder or attention deficit hyperactivity disorder or moderate or severe difficulty with emotions, concentration, or getting along with others. However, important caveats regarding NSCH data are that the assessment of family structure were subjectively based on respondents' self-reports and single-item indicators were used to assess children's physical and mental health (Bramlett & Blumberg, 2007).

With data from the 1999 and 2002 waves of the National Survey of America's Families (NSAF), Conway and Li (2012) examined relationships between diverse classifications of family structure with the physical, behavioral, and educational well-being of children between ages 6 and 17 years old. Similar to prior investigators, these authors found better child outcomes in families containing two biological adoptive/adoptive married parents compared to all nontraditional family structures, with children in “grandparent families” having the worst health. However, it is uncertain how “grandparent families” were defined operationally and the NSAF behavioral outcome measures relied on by Conway and Li (2012) are not well validated.

In a rare longitudinal study, Goulette et al. (2016) compared the self-reported antisocial and problematic behaviors of youth raised by CGMs, biological mothers, and other female caregivers from adolescence (10–17) to young adulthood (18–26) using data from Waves I, II, and III of the National Longitudinal Study of Adolescent Health (Add Health). Caregiver type was not found to significantly influence the number of antisocial acts reported by respondents over time, yet respondents raised by a CGM reported significantly lower rates of violence than those raised by biological mothers. However, it is critical to note that these self-reported measures were of questionable validity and did not assess mental health status per se.

Apart from the work of Goulette et al. (2016), the above referenced studies conducted with nationally representative samples suggest that CG experience greater emotional and behavioral difficulties in comparison to age peers living in more traditional family structures. Nevertheless, each of these studies had important limitations that are addressed in the present study. For example, three of the studies (Bramlett & Blumberg, 2007; Conway & Li, 2012; xlines are disconnected here xxxxxxxxx

Smith & Palmieri, 2007) involved children of such wide age ranges (from ages 4 to 17) that the extent of emotional and behavioral difficulties within narrower age groups may have been masked. Although, Conway and Li (2012) did compare younger and older children across diverse family structures, their findings were likely compromised by the fact that the NSAF data set used different indices of behavioral/emotional problems for children of ages 6–11 vs. 12–17.

1.3. Specific aims and hypotheses

In the present study, we specifically focus on reports from caregivers of primary school children between ages 4–12 because it is critical to screen children at early ages for mental health difficulties in order to prevent childhood problems from developing into severe psychiatric disorders (Stone, Otten, Engels, Vermulst, & Janssens, 2010). In addition, adolescents and younger children differ in the types of situations that create negative experiences for them, with primary school children experiencing a greater proportion of negative events tied to family members and their immediate surroundings than do older children who are in later adolescence (Lin, 2018; Spear, 2000). From a methodological point of view, there are also concerns that caregivers are less aware of how older adolescents feel and behave due to the inability to observe a full range of their behaviors (Langton & Berger, 2011).

With the exception of Smith and Palmieri (2007) the studies referenced above did not define specific inclusion or exclusion criteria regarding grandparent only family structure or screen accordingly. This is concerning because grandparent only families vary considerably in terms of how much biological parents are involved with the grandchild's life. For example, some birth parents may voluntarily arrange for their children to live with grandparents due to insufficient parental resources, because the birth parent is too young, military deployment, job or school responsibilities, and physical illness or injury (Vandivere et al., 2012). In the present study, we include only those grandparent headed families where grandchildren were being cared for under circumstances where neither of their birth parents either lived in the home or contributed meaningfully to child care. This distinction is crucial in view of longitudinal research showing that taking on higher levels of care responsibility was differentially associated with increased stress, intra-family strain, and worsened family functioning reported by grandmothers in differing types of caregiving roles (Musil et al., 2011).

Three of the studies referenced above (Bramlett & Blumberg, 2007; Conway & Li, 2012; Goulette et al., 2016) were further limited by the non-standardized measures of children's socio- emotional well-being present within the NSCH, NSAF, and Add Health data sets. Most notably, none of the measures in these three data sets were capable of assessing specific dimensions of children's psychological difficulties (e.g., internalizing and externalizing symptoms), had scoring procedures indicative of clinical-level disorders, or were validated for use with custodial grandparents. We offset these limitations in the present study by following Smith and Palmieri's (2007) approach of using the highly established SDQ which assesses multiple dimensions of children's psychological well-being and allows for the computation of scores that are indicative of children's serious psychological difficulties (Goodman, 1997, 2001).

In sum, the present study adds to existing knowledge regarding the risk of psychological difficulties among CG by comparing mean SDQ scores across two distinct samples of CG between ages 4 to 12 to sample children of the same ages from the 2004 National Health Interview Survey (NHIS) who either resided in households with no biological parents, one parent, or both parents. One sample of CG was from a nationwide survey study of CGMs (Smith, Palmieri, Hancock, & Richardson, 2008), whereas the second was recruited for a multi-state randomized clinical trial (RCT) comparing psychoeducational interventions delivered to their CGMs (Smith et al., 2018). By comparing these two samples of CG to their age peers in the general 2004 NHIS sample child population we attempt to cross validate and extend the findings of Smith and Palmieri (2007) who found CG between ages 2 to 17 to have more psychological difficulties as reported by their CGMs than the general sample population in the 2001 NHIS. This is an important aim because Smith and Palmieri's initial findings may have been limited to either their unique sample of CG, to children between ages 2 to 17, or to comparisons involving sample children from the earlier 2001 NHIS.

We hypothesize both CG samples in the present study will be rated as faring worse than their age peers in the general 2004 NHIS overall child sample, while having a similar degree of psychological difficulties as those children in the 2004 NHIS data set who lived in no-parent households. This expectation is based on the premise that CG are a unique subset of all children living in a household with no biological parents present. We additionally hypothesize that the degree of psychological difficulties reported for children will decrease as the number of parents in the household increases.

Another aim of the present study is to investigate the main effects of children's gender and race on psychological difficulties as indexed by the SDQ. We hypothesize that boys will be reported by their caregivers to have higher levels of externalizing symptoms than girls, whereas girls will be reported to have higher levels of internalizing symptoms than boys. This is consistent with studies showing that externalizing problems are seen more often in boys, while internalizing problems are more common in girls (Zahn-Waxler, Shirtcliff, & Marceau, 2008). Moreover, because these gender differences in disorder type reflect the extremes of normative characteristics of boys and girls (Zahn-Waxler et al., 2008), we expect to find them across all five samples.

Investigating race differences is important because the relationship between family structure and child well-being is more pronounced for White than for either Black or Hispanic children. It has been asserted that living outside of a two-biological parent family may be less stigmatizing and less stressful for minority children than for White children (Fomby & Cherlin, 2007). That turbulence in early childhood has a greater impact on subsequent behavior problems for White than for minority children may be due the fact that extrafamilial members such as grandparents play a more prominent role in parenting children within minority families, thereby minimizing the loss of a caregiver following family disruption (Womack, Taraban, Shaw, Wilson, & Dishion, 2018).

Our final aim is to calculate sample specific banded scores for each SDQ subscale as stipulated by Goodman (2001) across all study samples. This is critical because statistically significant differences in mean scores on screening scales like the SDQ do not necessarily indicate corresponding differences in risk for actual psychopathology (A. Goodman & Goodman, 2009). By computing SDQ sample specific banded scores we can estimate the percentage of children within each sample who reach a threshold indicative of “caseness”, which refers to children with clinically significant emotional or behavioral difficulties (Mark & Buck, 2006).

2. Method

2.1. Study samples

The survey sample data for the current study (n = 509) were drawn from 733 CGM recruited from 2002 to 2004 across 48 states through both convenience (e.g., support groups) and population-based (letters mailed at random) methods for a telephone interview study of stress and coping among custodial grandparents (Smith & Palmieri, 2007). Eligibility criteria were that CGMs had been caring for a CG between ages 4–17 for at least three consecutive months (with both birth parents absent from the household and reported by the CGM as not contributing in any meaningful way to CG care), and self-identified as either Black or White. If a CGM cared for multiple grandchildren, then a target CG was selected by the most recent birthday. Only data regarding target CG of ages 4–12 (n = 509) are used in the present study.

The RCT data on CG were drawn from a sample of 343 CGM recruited from 2011 to 2013 by convenience (e.g., media announcements, schools, service providers) across multiple states for a RCT study described to prospective participants as providing “information to help grandmothers get through the difficult job of caring for grandchildren in changing times” (Smith et al., 2018). Inclusion criteria were that CGMs provided care to a CG between ages 4 to 12 for at least 3 months in absence of the birth parents either residing in the home or

contributing in any meaningful way to CG care, were fluent in English, could attend 10 two hour group sessions, had not received any prior interventions like those under investigation, and self-identified as White, Black, or Hispanic. If a CGM cared for multiple CG of ages 4 to 12, then a target CG was identified as the one perceived as being the "most difficult" to care for. Of the 343 target CG, full SDQ baseline data were available for 323 in the present study.

Data regarding the three NHIS 2004 samples of U.S. children in the present study of ages 4 to 12 were obtained from the NHIS Web site http://www.cdc.gov/nchs/nhis.htm. The NHIS is a multi-purpose health survey conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention, and is the principal source of information on the health of the civilian, noninstitutionalized, household population in the United States. From each family in the NHIS, one sample adult and one sample child (if any children under age 18 are present) were randomly selected, and information on each was collected with the Sample Child Core and the Sample Adult Core questionnaires. For sample children, information was provided by a knowledgeable adult (KA) family member (18 years of age or over) residing in the household. Both the black and Hispanic populations were oversampled to allow for more precise estimation of health in these growing minority populations.

The final 2004 NHIS sample consisted of 36,579 households, which yielded 94,460 persons in 37,466 families. The final sample for the Sample Child component was 12,424 children under age 18, with a response rate of 79.4%. Although more recent NHIS data are available regarding a brief six item version of SDQ, the 2004 survey data were selected for the present study because 2004 is the final year that the 25-item version of the SDQ was used in the NHIS. A full description of the 2004 survey design and general characteristics of the corresponding data is available at http://www.cdc.gov/nchs/r&d/rdc.htm.

Of the 12,424 children sampled in the 2004 NHIS data set, only those between ages 4 to 12 with full SDQ data ($n = 5554$) were included in the present study and divided into three subsamples. The no parent (NP) sample contains 184 children who were reported to reside in a household with neither biological parent present at the time of the interview. Of these children, the KA was either a grandparent (66.3%), another relative (e.g., aunt/uncle) other than a biological parent (11.4%), a non-biological parent such as adoptive or step-parent (10.9%), a legal guardian or foster parent (7.0%), or other (4.3%). The one parent (OP) sample contains 1618 children reported to reside in a household with either a mother only ($n = 1402$) or a father only ($n = 216$). A Multivariate Analysis of Variance (MANOVA) comparing these two groups across all SDQ subscales was not significant ($F(5,1612) = 1.89$, $p < .09$), thus resulting in the decision to collapse them into a single NHIS OP sample irrespective of the parent's gender. Of these 1618 children, the KA was either a parent (88.0%), a grandparent (8.4%), a relative other than parent or grandparent (2.1%), a non-relative (1.1%), other (0.4%), or unknown (0.1%). The both parent (BP) sample contains 3752 children who were reported to be residing in a household with both birth parents present. Of these children, the KA was either a parent (98.1%), a grandparent (0.9%), a relative other than parent or grandparent (0.7%), a non-relative (0.2%), or other (0.2%).

2.2. Background characteristics of informants and target children

For the sake of convenience, the individuals reporting on the target child across all five samples of the present study are hereafter referred to as "informants". Self-reported demographic information pertaining to the sample children and their informants within each of the five samples is shown in Table 1. In those instances where original response alternatives differed by sample type (e.g., for education) recoding was done to make them comparable.

There were statistically significant differences between samples regarding informants' age, race, education, gender, as well as target child age. While the average age for both CGM samples and the NHIS NP sample informants was in the 50's, the average age for the NHIS OP and BP informants was in the late 30's. Although both CGM samples, were almost evenly split between Whites and non- Whites, White informants were predominant across all three NHIS samples. Generally speaking, both samples of CGM s had overall higher levels of education than informants from the NHIS samples. The mean age for target children in the NHIS No Parent Sample ($M = 8.8$ yrs.) was significantly higher than that for all other sample types, with the latter not differing significantly from one another regarding target child age. Inspection of target child age groupings by sample type in Table 1 further reveals that approximately one third of children in each sample type were in the respective age groups of 4–6, 7–9, and 11–12, with the exception of the NHIS No Parent sample where about only one fourth of children were in the 4–6 and 7–9 groups and nearly 50% were between ages 10–12.

By design all CGM informants were female, whereas small percentages of male informants were in all three NHIS samples. There were no statistically significant between sample differences regarding target child gender. The most frequently reported reasons for care being provided by CGMs were parents' substance abuse, physical or emotional abuse toward the child, child abandonment, parents' incarceration, and parents' mental health problems.

2.3. Measures

Data obtained from all samples in the present study were collected by trained interviewers over the telephone. Written informed consent was obtained from all CGM participants in both the survey and RCT samples.

The U.S. parent version of the SDQ (Goodman, 1997, 2001) was used to assess children's psychological strengths and difficulties. The SDQ contains 25 items divided equally among five scales: emotional symptoms, conduct problems, hyperactivity/inattention, peer problems, and prosocial behavior. Informants rated each item concerning the target child along a 3-point scale from 0, not true, to 2, certainly true. Summed scores on each subscale ranged from 0 to 10. A total difficulties score was also obtained by summing all subscales except for prosocial, with a range of 0 to 40. Key psychometric properties of the SDQ were previously found to be strong with NHIS samples (Bourdon, Goodman, Rae, Simpson, & Koretz, 2005), CG (Palmieri & Smith, 2007), and children between ages 4–12 in general (Stone et al., 2010).

We used the SDQ banded-scoring procedure (R. Goodman, 1997) by dividing the scores for the total difficulties scale and each subscale into two bands: (a) the highest 10% of scores representing "high difficulties" (or abnormal), and (b) the next 10% representing "medium difficulties" (or borderline). The remaining 80% of children are considered to be within the "normal" range. Banding was reversed for the prosocial scale. High scores are desirable for the prosocial scale; therefore, thus labeling of the bands was reversed. The percentage of children correctly identified by the SDQ as having a disorder is high, as is the percentage of children correctly identified by the SDQ as not having a disorder (see for review Stone et al., 2010). The NHIS SDQ data have been deemed "the best available resource for characterizing youths with serious emotional disturbances" (Mark & Buck, 2006). Full SDQ scoring procedures are available at www.sdqinfo.com.

2.4. Data analysis plan

A multivariate analysis of covariance (MANCOVA) was performed using SPSS v. 24 general linear modeling procedure to examine simultaneously the main effects of sample type, race, and target child gender on all SDQ subscales, with informant age and educational attainment as well as target child age included as covariates. Given that missing data for informants' age (0.11%) and education (0.27%) were infrequent and unsystematic, the mean value for each covariate across sample type was substituted. ES were calculated online to obtain Cohen's d for groups with different sample sizes (https://www.psychometrica.de/effect_size.html).

Table 1
Demographics by sample type.

	CGM survey		CGM RCT		NHIS NP		NHIS OP		NHIS BP		Test statistic
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	
Informant age (*M* | *SD*)	56.0 | 8.0		58.4 | 8.0		52.4 | 12.9		37.1 | 10.4		37.6 | 7.9		$F(4,6381) = 981.28^{***}$
Informant gender											$\chi 2(4) = 541.38^{***}$
Female	509	100.0%	323	100.0%	146	79.3%	1340	82.8%	2417	64.4%	
Informant race											$\chi 2(4) = 590.60^{***}$
White	257	50.5%	146	45.2%	115	62.5%	1031	63.7%	3115	83.0%	
Non-white	252	49.5%	177	54.8%	69	37.5%	587	36.3%	637	17.0%	
Informant education											$\chi 2(16) = 214.73^{***}$
< High school	85	16.7%	45	13.9%	66	35.9%	396	24.5%	680	18.1%	
High school graduate	139	27.3%	61	18.9%	59	32.1%	466	28.8%	938	56.4%	
Some college	176	34.8%	153	47.4%	44	23.9%	532	32.9%	1131	20.1%	
College graduate	70	13.8%	43	13.3%	7	38.0%	172	10.6%	665	17.7%	
Graduate/Professional degree	39	7.7%	21	5.1%	8	4.3%	52	3.2%	338	9.0%	
Child age (*M* | *SD*)	7.8 | 2.5		7.9 | 2.5		8.8 | 2.5		8.0 | 2.6		7.9 | 2.6		$F(4,6381) = 4.87^{***}$
Child age grouping											$2(8) = 17.78^{*}$
4–6 years	183	36.0%	115	35.6%	47	25.5%	538	33.3%	1303	34.7%	
7–9 years	163	32.0%	101	31.3%	50	27.2%	523	32.3%	1202	32.0%	
10–12 years	163	32.0%	107	33.1%	87	47.3%	557	34.4%	1247	33.2%	
Child gender											$2(4) = 9.07$
Female	272	53.4%	158	48.9%	99	53.8%	780	48.0%	1776	47.3%	
Reasons for care											
Child abandonment	139	27.3%	77	23.8%							
Parent substance abuse	262	51.5%	156	48.3%							
Physical/Emotional abuse	143	28.1%	127	39.3%							
Parent incarceration	229	45.0%	91	28.2%							
Removal from parent	163	32.0%									
Parent HIV/AIDS	4	0.8%									
Parent mental illness	150	29.5%	68	21.1%							
Other health problems	78	15.3%	13	4.0%							
Parent teen pregnancy	89	17.5%	25	7.7%							
Death of one parent	30	5.9%									
Sample size (*N*)	509		323		184		1618		3752		

Note: Reasons for care: Grandmothers were asked to report any and all reasons for assuming care of the target grandchild that applied to their situation.
* $p < .05$.
*** $p < .001$.

3. Results

The multivariate tests for main effects for sample type were statistically significant (F (20, 25,448) = 33.29, $p < .001$), race (F (5, 6359) = 12.67, $p < .001$), and child gender (F (5, 6359) = 27.29, $p < .001$). Because the multivariate tests were also statistically significant for the covariates of informant age (F (5,6359) = 3.49, $p < .005$), informant education (F(5,6359) = 28.59, $p < .001$), and target child age ($F_{(5,6359)} = 26.59$, $p < .001$), mean scores adjusted for these covariates are reported in all cases.

Tests of the between-subjects effects for sample type were statistically significant across all SDQ subscales ($p < .001$). Table 2 summarizes the mean comparisons by sample type for each SDQ subscale and the corresponding effect sizes (ES). A general pattern is seen where CG from the RCT sample were reported to have the worst mean scores on all SDQ scales (except for prosocial), with their CG peers from the RCT sample following closely behind. Overall, the mean SDQ scores of children from the NHIS samples were considerably better than those of both CG samples. Across the NHIS samples, SDQ scores were generally the best for children living with both parents and the worst for those with no parents in the household. All pairwise comparisons reported in Table 2 were statistically significant ($p < .001$), with the exception of no significant differences found between the CG survey and RCT samples on either the peer or prosocial scales.

The ES for the comparisons between the CG survey and RCT samples were all in the small range (0.06–0.34). The comparisons of the two CG samples with the NHIS NP sample ranged from 0.34–1.00, with ES for comparisons between the GC survey and NHIS NP samples being generally lower (0.39–0.65) than those for comparisons between the RCT and NHIS NP samples (0.34–1.00). The ES for mean differences between the CG survey and NHIS OP samples were medium to large in magnitude (0.56–0.81), whereas those for mean differences between CG RCT and NHIS OP samples were generally larger (0.55–1.15). While the majority of ES for the comparisons between the CG survey sample and the NHIS BP sample were in the medium to large range (0.46–0.83), the majority of the ES for the comparisons between the CG RCT and NHIS BP samples were generally larger (0.42–1.11). Generally speaking, ES associated with comparisons involving the prosocial scale were lower than the corresponding ES involving the difficulties oriented SDQ scales. When the ES associated with the prosocial subscale are disregarded, then all other statistically significant mean comparisons of the two CG samples with all three NHIS samples have ES that are of medium magnitude and higher. Generally speaking, these ES indicate that sample differences across SDQ difficulty-oriented subscales were reasonably robust in nature, speaking to the *importance* and *meaningfulness* of such differences rather than to their magnitude per se (Cohen, Cohen, West, & Aiken, 2003; Tabachnick & Fidell, 1996).

Tests of the between-subjects effects involving target child gender were statistically significant for all SDQ subscales ($p < .001$), except for emotional symptoms. These comparisons are summarized in Table 3 which shows the adjusted mean scores for each SDQ scale by sample type and child gender, and the corresponding F-tests. Regardless of sample type, scores for males were significantly worse on all SDQ scales other than emotional symptoms. However, the absolute magnitude of these gender differences are generally small irrespective of sample type.

As shown in Table 4, tests of the overall between-subjects effects for informant's race were statistically significant for the SDQ emotional symptoms ($p < .001$), hyperactive/inattention ($p < .05$), and total

Table 2
Mean comparisons by sample type and effect sizes.

SDQ subscale	Adjusted means and standard errors										Effect sizes for pairwise comparisons	
	Custodial grandfamilies				NHIS families							
	S (n = 509)		RCT (*n* = 323)		NP (*n* = 184)		OP (*n* = 1618)		BP (*n* = 37SS52)		S vs RCT	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*ES*	95% CI
Internalizing												
Peer	2.5	0.08	2.7	0.10	1.7	0.13	1.5	0.04	1.3	0.04	0.11[a]	−0.03–0.25
Emotional	2.7	0.10	3.3	0.12	1.7	0.15	1.5	0.05	1.2	0.04	0.27	0.13–0.41
Externalizing												
Hyperactivity	4.7	0.13	5.7	0.16	3.5	0.20	3.1	0.07	2.7	0.06	0.34	0.20–0.48
Conduct	2.7	0.09	3.2	0.11	1.6	0.13	1.4	0.05	1.0	0.04	0.25	0.11–0.39
Prosocial	7.8	0.08	7.9	0.10	8.5	0.13	8.8	0.04	8.9	0.04	0.06[a]	−0.08–0.20
Total difficulties	13	0.29	14.9	0.36	8.4	0.46	7.5	0.16	6.2	0.13	0.34	0.20–0.48

SDQ subscale	Effect sizes for pairwise comparisons											
	S vs NP		S vs OP		S vs BP		RCT vs NP		RCT vs OP		RCT vs BP	
	ES	95% CI	*ES*	95% CI	*ES*	95% CI	*ES*	95% CI	*ES*	95% CI	*ES*	95% CI
Internalizing												
Peer	0.45	0.28–0.62	0.60	0.50–0.70	0.50	0.41–0.60	0.56	0.38–0.74	0.73	0.61–0.85	0.58	0.47–0.70
Emotional	0.46	0.28–0.63	0.58	0.48–0.68	0.62	0.52–0.71	0.76	0.57–0.94	0.88	0.76–1.01	0.87	0.75–0.98
Externalizing												
Hyperactivity	0.42	0.25–0.59	0.56	0.46–0.66	0.56	0.46–0.65	0.78	0.59–0.97	0.91	0.79–1.03	0.83	0.71–0.94
Conduct	0.56	0.39–0.73	0.65	0.54–0.75	0.71	0.61–0.80	0.84	0.65–1.03	0.90	0.78–1.02	0.91	0.80–1.03
Prosocial	0.39	0.22–0.56	0.60	0.50–0.70	0.46	0.37–0.56	0.34	0.15–0.52	0.55	0.43–0.67	0.42	0.30–0.53
Total difficulties	0.65	0.48–0.82	0.81	0.70–0.91	0.83	0.74–0.93	1.00	0.81–1.19	1.15	1.02–1.27	1.11	0.99–1.22

[a] Sample Type pairwise *t*-test comparison not statistically significant; All other pairwise *t*-test comparisons are statistically significant.

Table 3
SDQ subscale means by sample type and target child gender.

SDQ Subscale	Adjusted means and standard errors											
	Survey		RCT		NHIS no parent		NHIS one parent		NHIS both parents		Gender main effect	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Internalizing												
Peer												
Male	2.7	0.11SS	3.0	0.13	1.9	0.18	1.6	0.06	1.3	0.05	2.1	0.05
Female	2.3	0.10	2.3	0.13	1.4	0.17	1.4	0.06	1.2	0.05	1.7	0.05
	$F(1,503) = 4.11^{**}$		$F(1,317) = 7.01^{**}$		$F(1,178) = 2.78$		$F(1,1612) = 10.35^{***}$		$F(1,3746) = 4.29^{*}$		$F(1,6363) = 28.01^{***}$	
Emotional												
Male	2.7	0.13	3.2	0.16	1.8	0.21	1.5	0.07	1.2	0.06	2.1	0.06
Female	2.7	0.12	3.5	0.16	1.6	0.20	1.5	0.07	1.3	0.06	2.1	0.06
	$F(1,503) = 0.16$		**$F(1,317) = 1.13$**		**$F(1,178) = 0.68$**		**$F(1,1612) = 0.16$**		**$F(1,3746) = 1.64$**		**$F(1,6363) = 0.40$**	
Externalizing												
Hyperactive												
Male	5.1	0.18	6.5	0.21	3.9	0.29	3.6	0.09	2.9	0.08	4.4	0.09
Female	4.3	0.17	4.9	0.22	3.1	0.27	2.6	0.10	2.4	0.08	3.5	0.08
	$F(1,503) = 9.35^{**}$		**$F(1,317) = 23.89^{***}$**		**$F(1,178) = 2.73$**		**$F(1,1612) = 54.38^{***}$**		**$F(1,3746) = 86.02^{***}$**		**$F(1,6363) = 73^{***}$**	
Conduct												
Male	3.0	0.12	3.6	0.14	1.8	0.19	1.6	0.06	1.1	0.05	2.2	0.06
Female	2.5	0.11	2.8	0.14	1.4	0.18	1.3	0.06	1.0	0.05	1.8	0.06
	$F(1,503) = 5.50^{*}$		**$F(1,317) = 6.17^{*}$**		**$F(1,178) = 2.83$**		**$F(1,1612) = 7.52^{**}$**		**$F(1,3746) = 13.10^{***}$**		**$F(1,6363) = 31.01^{***}$**	
Prosocial												
Male	7.6	0.11	7.4	0.13	8.1	0.18	8.5	0.06	8.8	0.05	8.1	0.06
Female	8.1	0.11	8.4	0.14	8.8	0.17	9.0	0.06	9.0	0.05	8.6	0.05
	$F(1,503) = 7.52^{**}$		**$F(1,317) = 21.53^{***}$**		**$F(1,178) = 9.47^{**}$**		**$F(1,1612) = 28.80^{***}$**		**$F(1,3746) = 27.83^{***}$**		**$F(1,6363) = 67.02^{***}$**	
Total difficulties												
Male	13.5	0.41	16.3	0.48	9.4	0.66	8.3	0.21	6.5	0.18	10.8	0.20
Female	11.9	0.38	13.6	0.50	7.5	0.62	6.7	0.22	5.9	0.19	9.1	0.19
	$F(1,503) = 5.49^{*}$		**$F(1,317) = 10.06^{**}$**		**$F(1,178) = 3.07$**		**$F(1,1612) = 20.65^{***}$**		**$F(1,3746) = 30.25^{***}$**		**$F(1,6363) = 42.60^{***}$**	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

difficulties ($p < .001$) scales only. Regardless of sample type, White informants reported significantly higher mean scores for children on the emotional symptoms scale than did non-White informants. Despite the statistically significant overall between-subjects effect on informant's race on the hyperactive/inattention scale ($p < .05$), none of the specific race comparisons across the five sample types were significant. Statistically significant mean differences by informant's race on the SDQ total difficulties scale were found only within the CG survey ($p < .05$) and CG RCT ($p < .05$) samples.

The results from the banded scores for sample-specific children with "high" (abnormal) and "medium" (borderline) difficulties are shown in Table 5, where all sample-specific banded scores are compared to the NHIS-BP sample. Score ranges associated with high and medium difficulties were uniformly higher for both the CG survey and CG RCT samples than they were for all three of the NHIS samples. When banded scores from the NHIS BP sample were applied to both CG samples, nearly 60% of CG fell within either the medium or high difficulties categories on all SDQ scales other than prosocial. Low scores on the latter scale, however, are not believed to be indicative of psychopathology (Goodman, 1997, 2001; Stone et al., 2011).

4. Discussion

We conducted the first study comparing the psychological difficulties of CG ages 4 to 12 as reported by CGM to those reported by adult informants for similarly aged children from diverse family structures sampled within the 2004 NHIS nationally representative data set. After controlling for differences by sample type in target child age as well as informant's chronological age and education, we found that both a sample of CGM recruited for a survey study and a sample of CGM recruited for a RCT reported significantly higher mean levels of GC difficulties across all six dimensions assessed by the prominent and psychometrically sound SDQ (Goodman & Goodman, 2009). In addition, as hypothesized, a pattern emerged whereby mean differences between all CG and sample children from the 2004 NHIS were magnified as the number of parents residing in a household increased from no parents present to both parents present. That this pattern was accompanied by substantial ES, rules out the possibility that the observed statistically significant differences were due solely to large sample sizes (Cohen, 1988). Indeed, the fact that multiple samples of children are compared regarding SDQ scores enhances the generalizability of concerns regarding the adjustment difficulties that many children raised by grandparents experience; this is underscored by the generally robust nature of such comparisons as reflected in ES associated with them, speaking to their meaningfulness and clinical importance.

Statistically significant differences were also found between the two GC samples on four of the six SDQ scales. However, given that CGMs from the RCT were recruited for a study designed to "help" their families, it is not surprising that they reported greater amounts of CG difficulties than the CGMs who participated in the more rudimentary survey study. Those CGM who seek help are not only more likely to be distressed themselves, but they are also likely to report behavioral problems within their CG (Daly & Glenwick, 2000). It is also possible, however, that differences between these two samples in how target CG was selected if multiple CG were in the home may have influenced our findings. Under these conditions, target CG in survey sample were selected at random, whereas target CG were selected by asking CGM to identify the CG she found most difficult to care for. Nevertheless, the differences we found here between the two CG samples involved ES which paled in comparison to the ES associated with the differences

Table 4
SDQ subscale means by sample type and race.

SDQ Subscale	Adjusted means and standard errors											
	Survey		RCT		NHIS no parent		NHIS one parent		NHIS both parents		Race main effect	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Internalizing												
Peer												
White	2.6	0.11	2.7	0.14	1.6	0.15	1.5	0.05	1.1	0.03	1.9	0.05
Non-white	2.4	0.11	2.7	0.13	1.7	0.19	1.5	0.07	1.4	0.06	1.9	0.06
	F(1,503) = 1.03		***F(1,317) = 0.16***		***F(1,178) = 0.10***		***F(1,1612) = 0.46***		***F(1,3746) = 14.97***		**F(1,6363) = 0.06**	
Emotional												
White	3.1	0.13	3.8	0.16	2.1	0.18	1.7	0.06	1.3	0.04	2.4	0.06
Non-white	2.3	0.13	2.8	0.15	1.4	0.23	1.4	0.08	1.2	0.08	1.8	0.07
	F(1,503) = 13.83		***F(1,317) = 14.82***		***F(1,178) = 4.13**		***F(1,1612) = 6.55**		***F(1,3746) = 2.26***		**F(1,6363) = 46.64***	
Externalizing												
Hyperactive												
White	5.0	0.17	6.0	0.22	3.7	0.24	3.0	0.08	2.6	0.05	4.1	0.08
Non-white	4.5	0.17	5.4	0.20	3.3	0.31	3.1	0.11	2.7	0.10	3.8	0.09
	F(1,503) = 3.71		***F(1,317) = 2.57***		***F(1,178) = 0.71***		***F(1,1612) = 0.73***		***F(1,3746) = 0.69***		**F(1,6363) = 4.38***	
Conduct												
White	2.8	0.11	3.4	0.15	1.7	0.16	1.3	0.05	1.0	0.03	2.0	0.05
Non-white	2.7	0.11	3.1	0.14	1.5	0.21	1.5	0.07	1.0	0.07	2.0	0.06
	F(1,503) = 0.09		***F(1,317) = 1.28***		***F(1,178) = 0.48***		***F(1,1612) = 2.57***		***F(1,3746) = 0.37***		**F(1,6363) = 1.51**	
Prosocial												
White	7.9	0.11	8.0	0.14	8.6	0.16	8.8	0.05	9.0	0.03	8.4	0.05
Non-white	7.7	0.11	7.8	0.13	8.4	0.20	8.7	0.07	8.8	0.07	8.3	0.06
	F(1,503) = 0.64		***F(1,317) = 0.21***		***F(1,178) = 0.28***		***F(1,1612) = 0.87***		***F(1,3746) = 9.46***		**F(1,6363) = 3.33**	
Total difficulties												
White	13.4	0.39	15.9	0.51	9.0	0.56	7.5	0.19	6.1	0.11	10.4	0.18
Non-white	11.9	0.40	14.0	0.47	7.8	0.72	7.5	0.25	6.3	0.23	9.5	0.21
	F(1,503) = 5.11*		**F(1,317) = 5.59***		**F(1,178) = 1.14**		**F(1,1612) = 0.03**		**F(1,3746) = 0.62**		**F(1,6363) = 11.49***	

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 5
Percentage of SDQ cut scores at 80th (borderline) and 90th (abnormal) percentiles across samples.

SDQ subscales	Sample-specific cut-off scores				NHIS both parent sample cut-off scores			
	80th Percentile		90th Percentile		80th Percentile		90th Percentile	
	Score range	%	Score range	%	Score range	%	Score range	%
Survey sample (N = 509)								
Peer	4	10.6	≥5	17	2	18.7	≥3	40.9
Emotional	5	7.9	≥6	13.4	2–3	26.7	≥4	30.5
Hyperactive	8	7.9	≥9	13.8	5	10.2	≥6	36.3
Conduct	5	6.3	> 6	14.1	2	18.5	> 3	43
Total difficulties	19–22	10.1	≥23	12.1	10–13	16.8	≥14	38.9
Prosocial	5	6.3	< 4	7.9	7	13.6	< 6	24.4
RCT Sample (N = 323)								
Peer	4	12.1	≥5	18.6	2	16.7	≥3	47.4
Emotional	5	9.6	> 6	18.6	2–3	27.6	> 4	39.9
Hyperactive	9	11.8	10	11.5	5	13	≥6	47.7
Conduct	5–6	13.9	≥7	10.8	2	18	≥3	52
Total Difficulties	22–25	10.0	≥26	11.1	10–13	18.5	≥14	49.2
Prosocial	5	6.5	≤4	6.8	7	13.3	≤6	23.2
NHIS NP sample (N = 184)								
Peer	3	9.2	> 4	16.8	2	16.8	> 3	26.1
Emotional	3–4	14.7	≥5	13.0	2–3	22.3	≥4	17.9
Hyperactive	6–7	9.8	≥8	15.8	5	10.3	≥6	22.3
Conduct	3	13.0	> 4	13.6	2	13.0	> 3	26.6
Total difficulties	13–18	12.4	> 19	10.3	10–13	17.4	> 14	18.5
Prosocial	7–6	11.6	≤5	5.4	7	7.6	≤6	17.9
NHIS OP Sample (N = 1618)								
Peer	3	9.3	≥4	12.8	2	19.6	≥3	22.1
Emotional	3	8.5	> 4	16.0	2–3	22.3	> 4	16.0
Hyperactive	5–6	13.2	≥7	13.4	5	7.2	≥6	19.5
Conduct	3	7.7	> 4	13.9	2	15.3	> 3	21.6
Total difficulties	12–17	11.6	≥18	10.2	10–13	12.3	≥14	16.9
Prosocial	7–6	13.2	< 5	6.1	7	8.3	< 6	10.9

between these two samples and sample children from the 2004 NHIS. In other words, the two CG samples were far more similar to one another than they were to all children of the same age sampled within the 2004 NHIS.

Surprisingly, the findings here do not fully support our hypothesis that the degree of psychological difficulties reported for CG would be similar to those reported for those NHIS children residing in homes with no parent present in view of prior evidence that living in a household with no biological parents present is associated with children's poor physical and mental health outcomes (Beal & Greiner, 2016; Bramlett & Blumberg, 2007; Vandivere et al., 2012; Ziol-Guest & Dunifon, 2014). This is especially puzzling given that 66.3% of the informants from the 2004 NHIS NP sample were the sample child's grandparent. One possible explanation is that the reasons why no parents were present in the NHIS households may have been quite different than the types of parental dysfunction or incapacity which typically lead to OOHC by custodial grandparents (Hayslip et al., 2017). For example, it is possible that a considerable number of 2004 NHIS NP households were the result of such factors as parents' military deployment, lack of financial resources, homelessness, or birth parents' very young age (Edwards, 2009; Vandivere et al., 2012; Washington et al., 2018). As such, the OOHC living arrangements for many of these children may have been only temporary and involved some degree of meaningful parent-child interaction in comparison to the anticipated permanent placements with minimal or no parental involvement that predominate within custodial grandfamilies. Thus, our findings point to the key unanswered question of whether the elevated behavior problems reported for CG "are the result of conditions precipitating placement with grandparents (e.g. child maltreatment, parental incarceration), disruption in attachment relationships with birth parents, or circumstances in the current caregiving environment (p. 2143 Kelley, Whitley, & Campos, 2011).

Our second aim was to investigate the main effects of children's gender and race. Regarding the effects of gender, we hypothesized that boys would be reported by informants across all five samples to have higher levels of externalizing symptoms than girls, whereas girls would be reported to have higher levels of internalizing symptoms. This hypothesis was fully supported with respect to externalizing symptoms, as boys within all five samples were reported to have statistically significant higher levels of conduct problems and hyperactivity/inattention than girls. On the other hand, our predictions regarding gender differences for internalizing symptoms were unsupported. Not only were there no significant differences between boys and girls within all five samples regarding emotional symptoms, boys were unexpectedly reported to have significantly higher levels of peer problems than girls across all samples except for the NHIS NP sample. Consistent with our hypotheses, and irrespective of sample type, boys were reported to also have significantly less prosocial behaviors and greater total difficulties (except for the NHIS NP sample) than girls.

Our unanticipated findings regarding gender differences may be due in part to normative developmental trajectories whereby early onset externalizing disorders (e.g., conduct disorder, attention deficit-hyperactivity disorder) show a marked male preponderance, whereas adolescent onset internalizing disorders (e.g., depression, anxiety) show a marked female preponderance (Zahn-Waxler et al., 2008). Early prenatal exposure to testosterone may account for slower maturation and greater disinhibition in boys, while increased estrogen during puberty may delay the ability of girls to recover from stress and leave them more susceptible to internalizing difficulties than boys (Spear, 2000; Zahn-Waxler et al., 2008). Our explicit focus on children ages 4–12 may have brought these developmental issues to the forefront. Along these lines, it is noteworthy that Simpson, Bloom, Cohen, Blumberg, and Bourdon (2005) found that 15–17 year old girls sampled in the 2003 NHIS were almost three times likely as 4–7 year old girls to have either "definite" or "severe" difficulties as measured by a single SDQ question. The present findings clearly point to the need for longitudinal studies on the etiology and progression of psychological difficulties by children's gender across all family structures. This need is especially true for custodial grandfamilies give that little is presently known about how the transition to care provided by custodial grandparents affects CG's socioemotional well- being.

Our observed main effects regarding race were modest, with statistically significant univariate effects along this variable occurring for only the emotional symptoms and total difficulties scales of the SDQ. Across all five samples, white informants reported significantly higher levels of children's emotional symptoms than did non-white informants. However, significant race differences regarding total difficulties were limited to white CGMs reporting greater total difficulties scores for CG than non-white grandmothers. That White CGMs reported more total difficulties than non-White CGMs might be attributable to the fact that White caregivers tend to experience greater negative affect and burden than non-Whites (Goodman & Silverstein, 2006; Pruchno, 1999) because they have fewer cultural expectations and reinforcements for assuming extended care of a GC than do non-Whites (Womack et al., 2018). In turn, such greater negative affectivity in White CGMs may either negatively bias their perceptions of CG behavior or increase the level of GC difficulties through poor parenting and negative role modeling (Smith et al., 2018). Our overall finding of relatively few race differences despite the fact that children of color are at risk for psychological difficulties because of their minority status, often lower SES, and limited access to mental health services suggests the need to further identify potential protective factors that may be effectively buffering them against the stressors of poverty, prejudice, and social isolation (Washington et al., 2018). Nevertheless, an important caveat is that our findings based on informants' race do not necessarily correspond to child race precisely.

Given that differences in mean SDQ scores between groups do not necessarily correspond to salient differences in mental health status, our final aim was to examine the SDQ banded scores which categorized children within all five samples into "medium" (borderline) and "high difficulties" (abnormal) groups based on their scores within at 80th and 90th percentiles, respectively. The high difficulties category is intended to signify those children who are likely to have serious mental health difficulties or "caseness" (Goodman, 1997). It is critical to note that Goodman (1997) has advocated for the calculation of sample specific banded scores, as was done here, because caseness "does not have a comparable meaning in different studies simply because those studies have employed the same cut-off "(pp. 584–585).

We found that for each SDQ scale, the sample specific cut scores for medium and high difficulties differed substantially by sample type, with the highest cut scores observed for both GC samples and the lowest for the NHIS BP sample. Indeed, Goodman's (1997) recommendation to use sample specific banded scores seems sensible in the present instance given that around 30–50% of all GC would fall into the high difficulties category if banded scores calculated for the NHIS BP sample were applied instead. However, Goodman (1997) further advised that until studies are conducted to precisely define cut-offs for different samples of children, the "medium difficulties" cut-offs can be applied to high risk samples while the "high difficulties" cut-offs can be used for low risk samples. The rationale for this recommendation is that the potential identification of false positives is not a major concern when high risk community samples, like CG, are being studied (Stone et al., 2010). Ideally, CG screened by the SDQ to fall within the "medium" or "high" difficulties ranges would be referred for thorough a diagnostic assessment.

As seen in Table 5, when sample specific SDQ banded scores in both the "medium" and "high" difficulty ranges are used to identify "caseness" in both CG samples, serious psychological disorder is evidenced in between 20 and 30% of CG on each SDQ scale except for prosocial (which does not indicate psychopathology per se). These percentages are remarkably similar to those reported by Kelley et al., (2011) who found that 31% of 230 CG ages 2 to 16 from a major metropolitan area in the Southeastern U.S. scored in the clinically elevated range for total

behavior problems on the Child Behavior Checklist (Achenbach, 2010), with 21.3% and 32.6% scoring in the elevated range for internalizing and externalizing behaviors, respectively. Importantly, these rates of psychological disorder are higher than the 3 to 18% rate that exists among the general population of children.

5. Limitations and directions for future research

There are several limitations associated with the present study which point to important directions for future research. One major shortcoming is that all data regarding children's psychological difficulties were obtained from a sole informant via the SDQ parent version. This is a concern because different informants such as parents, teachers, or neighbors view children's behaviors within varying contexts that may influence their reporting in different ways (Merchant, Hayslip, Smith, & Estrada, 2018; Stone et al., 2010). This concern, however, is diminished by findings from two smaller scale studies where teachers reported higher levels of overall psychological difficulties for CG than for their school peers. Using the teacher version of the CBCL (Achenbach, 1991), Edwards (2006) found 27 teachers to perceive significantly more CG in Grades 3–5 as demonstrating overall psychopathology than a carefully matched comparison group of children raised by parents. More recently, Pilkauskas and Dunifon (2016) found with several standardized measures that teachers reported CG (*M* age = 9.5 yrs.) to be significantly less cooperative, display more oppositional behavior, have more cognitive and/or inattention problems, and to be less likely to concentrate than schoolmates. Yet, to date, no investigators have obtained data from multiple informants (e.g., caregivers, teachers, child self-reports) regarding CG psychological difficulties and compared them to actual psychiatric diagnoses, which is an important direction for future research.

Although we were able to control for differences in informant's age and education as well as target child age, the present study was limited by our inability to control for other key variables that may have influenced the findings. In particular, due to extensive missing data on household income across all five samples, we were unable to control for the potential confounding effect of this variable which is a key socioeconomic risk factor for psychopathology in addition to education (Korous, Causadias, Bradley, & Luthar, 2018). This is an important concern given that nearly one third of CG live below the poverty line (Baker & Mutchler, 2010), and the corresponding economic hardships are likely to result in high stress family environments that negatively impact CC physical and emotional well-being (Hayslip et al., 2017; Kelley et al., 2011). Controlling for differences in children's physical health across samples would likewise have been appropriate given that CG fare significant worse than children in other family structures on physical health indices (Bramlett & Blumberg, 2007; Ziol-Guest & Dunifon, 2014); CG are at disproportionate risk of health insecurity (Baker & Mutchler, 2010); and children's poorer physical health is associated with higher SDQ scores (Stone et al., 2010). Likewise, we were unable to account for the potential disruptive influence of the parent on CG's mental health; varying degrees of parental involvement and contact may be associated with greater grandchild difficulties (Hayslip & Kaminski, 2006).

Another shortcoming is that the informants for target CG were entirely their CGM, whereas those for children across the 2004 NHIS samples varied by informant's gender and relationship to the target child. In turn, this points to the knotty methodological question of what constitutes the best family comparison groups when attempting to draw conclusions about the psychological well-being of CG (Pilkauskas & Dunifon, 2016). Difficulty in identifying meaningful comparison groups is problematic due to differences and similarities among family structures beyond socioeconomic status itself, including such contextual factors as household structure, caregiver characteristics, relationship quality, and available social support. In the present study, we dealt with this quandary by selecting comparison family structures that varied in terms of the number of parents present given that stability and continuity of attachment relationships are critical to the well-being of children, with the number of parents in the home serving as a proxy for such stability and continuity (Vandivere et al., 2012). An important next step is to investigate which specific contextual features within custodial grandfamilies have the greatest impact on the overall socioemotional well-being of CG, especially those factors which are responsive to interventions.

Our comparisons involving race were limited to comparing Whites and non- Whites, which may overlook important distinctions that exist between different racial and ethnic populations within the overly broad non-White category. However, because of the relatively small amount of racial diversity present within both the custodial grandfamily and NHIS samples, there was insufficient power to further divide the non-White samples into meaningfully subgroups (e.g., African American, Hispanic, and Asian). This problem is especially true for such family structures as custodial grandfamilies and other no parent families where their population prevalence is very low to begin with (Bramlett & Blumberg, 2007). In addition, our reliance on informants' self-reported race does not permit generalizations regarding race of the target child.

Finally, although our explicit focus was on children between ages 4 to 12, the present findings do not necessarily generalize to children of younger and older ages. Most importantly, future studies are needed which compare longitudinally the psychological adjustment of CGC from birth onward to similarly aged peers raised within diverse family structures. Future longitudinal studies would also reduce the possibility of cohort effects which may have influenced the present cross-sectional findings.

6. Conclusions

The results of this study are consistent with prior investigations pointing to elevated risk of psychological difficulties within CG which have included comparisons of CG to nationally representative samples (Bramlett & Blumberg, 2007; Conway & Li, 2012; Goulette et al., 2016; Smith & Palmieri, 2007), smaller scale regional studies (Dunifon & Kopko, 2011; Edwards, 2009; Kelley et al., 2011; Pittman, 2007), diverse family types sampled across large U.S. cities (Pilkauskas & Dunifon, 2016), as well as clinical reports (Harinder, Weist, & Shafer, 1999). Our findings add to this knowledge base uniquely by showing that two different national samples of CG between ages 4–12 were reported to have more psychological difficulties in comparison to similarly aged children from three different family structures sampled in the 2004 NHIS irrespective of race and gender. Collectively, these studies point to the need for improving the custodial grandfamily's access to appropriate mental health services in order to best screen, diagnose, and treat the emotional and behavioral problems of CG. Fortunately, the recent Supporting Grandparents Raising Grandchildren Act passed by the U.S. Congress (GovTrack.us, 2019) calls for the establishment of a Federal Advisory Council to make recommendations that would help grandparents and other older relatives better meet the health, educational, nutritional, and other needs of the children in their care, as well as maintain their own physical and mental health and emotional well-being.

Notwithstanding our explicit focus on CG psychological difficulties, it is critical to take note of substantial evidence that even though children receiving OOHC in kinship placements are more vulnerable than the general child population, they nonetheless have fewer behavioral problems than do children in foster care (see, for review, Washington et al., 2018). Thus, kinship care may provide a protective effect against psychological difficulties in those children receiving OHHC because family members like CGM are likely to be more invested than non-relative caregivers and may offer greater stability to the children in their care by maintaining key family, cultural, and

community ties (Font, 2014). Hence, we recommend that researchers and practitioners assume a resilience perspective with the goal of identifying which risk (e.g., early life adversity, caregiver stress, poverty) and protective factors (e.g., community resources, close interpersonal ties, self-regulation skills) exert the greatest impact on multiple developmental outcomes in the lives of CG (Hayslip & Smith, 2013).

In addition to psychological difficulties, other key outcomes for future deliberations of CG well-being include physical health and safety, cognitive development and education, psychological/emotional development, and social development (Lippman, Moore, & McIntosh, 2011). Until then, our findings highlight the need for clinicians to not only identify those children raised by grandparents who are at substantial risk for psychological difficulties, but to also provide interventions that are specifically targeted to the unique needs of custodial grandfamilies (McLaughlin, Ryder, & Taylor, 2017).

Funding

This work was supported by grant R01MH66851 from the National Institute of Mental Health awarded to the first author, and by grant R01NR012256 from the National Institute on Nursing Research awarded to the first and second authors.

Declaration of Competing Interest

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgements

We wish to thank NHIS support staff for their invaluable help with understanding and managing the 2004 NHIS data sets used in the present study.

References

Achenbach, T. (2010). *Manual for the child behavior checklist/4–18 and 1991 profile.* Burlington: University of Vermont, Department of Psychiatry.

Baker, L. A., & Mutchler, J. E. (2010). Poverty and material hardship in grandparent-headed households. *Journal of Marriage and Family, 72*(4), 947–962. https://doi.org/10.1111/j.1741-3737.2010.00741.x.

Beal, S., & Greiner, M. (2016). Children in nonparental care: Health and social risks. *Pediatric Research, 72*.

Billing, A., Ehrle Macomber, J., & Kortenkamp, K. (2002). *Children cared for by relatives - what do we know about their well-being? Washington, DC 20037*.

Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The strengths and difficulties questionnaire: U.S. normative data and psychometric properties. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(6), 557–564. https://doi.org/10.1097/01.chi.0000159157.57075.c8.

Bramlett, M. D., & Blumberg, S. J. (2007). Family structure and children's physical and mental health. *Health Affairs, 26*(2), 549–558. https://doi.org/10.1377/hlthaff.26.2.549.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Lawrence Erlbaum Associates.

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Lawrence Erlbaum.

Conley, C. S., Caldwell, M. S., Flynn, M., Dupre, A. J., & Rudolph, K. D. (2004). Parenting and mental health. *Handbook of parenting: Theory and research for practice* (pp. 276–295). Thousand Oaks, CA: SAGE Publications Inc. https://doi.org/10.4135/9781848608160.n17.

Conway, K. S., & Li, M. (2012). Family structure and child outcomes: A high definition, wide angle "snapshot.". *Review of Economics of the Household, 10*(3), 345–374. https://doi.org/10.1007/s11150-011-9121-x.

Daly, S. L., & Glenwick, D. S. (2000). Personal adjustment and perceptions of grandchild behavior in custodial grandmothers. *Journal of Clinical Child and Adolescent Psychology, 29*(1), 108–118. https://doi.org/10.1207/S15374424jccp2901_11.

Dunifon, R., & Kopko, K. (2011). Relationship quality and parenting among grandparent caregivers and teens. *Meeting of the Brookdale foundation relativs as parents conference* (Denver, CO).

Edwards, O. W. (2009). Empirical investigation of the psychosocial functioning of children raised by grandparents. *Journal of Applied School Psychology*. https://doi.org/10.1080/15377900802484653.

Fomby, P., & Cherlin, A. J. (2007). Family instability and child well-being. *American Sociological Review, 72*, 181–204.

Font, S. A. (2014). Kinship and nonrelative foster care: The effect of placement type on child well-being. *Child Development, 85*(5), 2074–2090. https://doi.org/10.1111/cdev.12241.

Goodman, A., & Goodman, R. (2009). Strengths and difficulties questionnaire as a dimensional measure of child mental health. *Journal of the American Academy of Child and Adolescent Psychiatry, 48*, 400–403. https://doi.org/10.1097/CHI.0b013e3181985068.

Goodman, C. C., & Silverstein, M. (2006). Grandmothers raising grandchildren: Ethnic and racial differences in well-being among custodial and coparenting families. *Journal of Family Issues, 27*, 1605–1626. https://doi.org/10.1177/0192513X06291435.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*(5) (581–58).

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 1337–1345.

Goulette, N. W., Evans, S. Z., & King, D. (2016). Exploring the behavior of juveniles and young adults raised by custodial grandmothers. *Children and Youth Services Review, 70*, 349–356. https://doi.org/10.1016/j.childyouth.2016.10.004.

GovTrack.Us (2019). S. 1091 – 115th Congress: Supporting Grandparents Raising Grandchildren Act. Retrieved from https://govtrack.us/congress/bills/115s1091.

Harinder, S., Weist, M., & Shafer, M. E. (1999). Demographic and clinical characteristics of emotionally disturbed children being raised by grandparents. *Psychiatric Services, 50*, 1496–1498.

Hayslip, B., Fruhauf, C., & Doolbin-MacNab, M. (2017). Grandparents raising grandchildren: What have we learned over the past decade. *The Gerontologist, 57*(6), 1196–2008.

Hayslip, B., & Kaminski, P. (2006). Custodial grandchildren. *Children's needs III: Understanding and addressing the needs of children* (pp. 771–782). Washington, DC: National Association of School Psychologists.

Hayslip, B., & Smith, G. C. (2013). *Resilient grandparent caregivers: A strengths-based perspective.* New York: Routledge.

Kelley, S. J., Whitley, D. M., & Campos, P. E. (2011). Behavior problems in children raised by grandmothers: The role of caregiver distress, family resources, and the home environment. *Children and Youth Services Review, 33*, 2138–2145. https://doi.org/10.1016/j.childyouth.2011.06.021.

Korous, K. M., Causadias, J. M., Bradley, R. H., & Luthar, S. S. (2018). Unpacking the link between socioeconomic status and behavior problems: A second-order meta-analysis. *Development and Psychopathology, 30*, 1889–1906. https://doi.org/10.1017/S0954579418001141.

Lippman, L. H., Moore, K. A., & McIntosh, H. (2011). Positive indicators of child well-being: A conceptual framework, measures, and methodological issues. *Applied Research in Quality of Life, 6*(4), 425–449. https://doi.org/10.1007/s11482-011-9138-6.

Mark, T., & Buck, J. (2006). Characteristics of US youths with serious emotional disturbance: Data from the national health interview survey. *Psychiatric Services, 57*, 1573–1578.

McLaughlin, B., Ryder, D., & Taylor, M. F. (2017). *Effectiveness of interventions for grandparent caregivers: A systematic review. marriage and family review.* https://doi.org/10.1080/01494929.2016.1177631.

Merchant, W., Hayslip, B., Smith, G., & Estrada, S. (2018). *Difficulties among custodial grandchildren: Whose view really matters?* (Boston, MA).

Musil, C. M., Gordon, N. L., Warner, C. B., Zauszniewski, J. A., Standing, T., & Wykle, M. (2011). Grandmothers and caregiving to grandchildren: Continuity, change, and outcomes over 24 months. *Gerontologist, 51*(1), 86–100. https://doi.org/10.1093/geront/gnq061.

Palmieri, P. A., & Smith, G. C. (2007). Examining the structural validity of the strengths and difficulties questionnaire (SDQ) in a U.S. sample of custodial grandmothers. *Psychological Assessment, 19*(2), 189–198. https://doi.org/10.1037/1040-3590.19.2.189.

Pilkauskas, N. V., & Dunifon, R. E. (2016). Understanding grandfamilies: Characteristics of grandparents, nonresident parents, and children. *Journal of Marriage and Family, 78*, 623–633. https://doi.org/10.1111/jomf.12291.

Pittman, L. D. (2007). Grandmothers' involvement among young adolescents growing up in poverty. *Journal of Research on Adolescence, 17*(1), 89–116. https://doi.org/10.1111/j.1532-7795.2007.00513.x.

Pruchno, R. (1999). Raising grandchildren: The experiences of black and white grandmothers. *The Gerontologist, 39*(2), 209–221.

Rubin, D. M., Downes, K. J., O'reilly, A. L. R., Mekonnen, R., Luan, X., & Localio, R. (2008). Impact of kinship care on behavioral well-being for children in out-of-home care one thousand three hundred nine children entering out-of-home care following a maltreatment report. *Archives of Pediatrics & Adolescent Medicine,* 162.

Sakai, C., Lin, H., & Flores, G. (2011). Health outcomes and family services in kinship care. *Archives of Pediatrics & Adolescent Medicine, 165*(2), 159–165. https://doi.org/10.1001/archpediatrics.2010.277.

Simpson, G. A., Bloom, B., Cohen, R. A., Blumberg, S., & Bourdon, K. H. (2005). U.S. children with emotional and behavioral difficulties: Data from the 2001, 2002, and 2003 national health interview surveys. *Advance Data,* (360), 1–16.

Smith, G. C., Hayslip, B., Hancock, G. R., Strieder, F. H., & Montoro-Rodriguez, J. (2018).

A randomized clinical trial of interventions for improving well-being in custodial grandfamilies. *Journal of Family Psychology, 32*(6), 816–827. https://doi.org/10.1037/fam0000457.

Smith, G. C., & Palmieri, P. A. (2007). Risk of psychological difficulties among children raised by custodial grandparents. *Psychiatric Services, 58*(10), 1303–1310.

Smith, G. C., Palmieri, P. A., Hancock, G. R., & Richardson, R. A. (2008). Custodial grandmothers' psychological distress, dysfunctional parenting, and grandchildren's adjustment. *The International Journal of Aging and Human Development, 67*(4), 327–357. https://doi.org/10.2190/AG.67.4.c.

Spear, L. P. (2000). The adolescent brain and age-related behavioral manifestations. *Neuroscience and Biobehavioral Reviews, 24*, 417–463. https://doi.org/10.1016/S0149- 7634(00) 00014.

Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review, 13*, 254–274. https://doi.org/10.1007/s10567-010-0071-2.

Tabachnick, B., & Fidell, L. (1996). *Using multivariate statistics.* New York: Harper Collins.

Vandivere, S., Yrausquin, A., Allen, T., Malm, K., & McKindon, A. (2012). Children in nonparental care: A review of the literature and analysis of data gaps. Retrieved from https://aspe.hhs.gov/basic-report/children-nonparental-care-review-literature-and-analysis- data-gaps.

Washington, T., Wrenn, A., Kaye, H., Priester, M. A., Colombo, G., Carter, K., & Coakley, T. (2018). Psychosocial factors and behavioral health outcomes among children in Foster and Kinship care: A systematic review. *Children and Youth Services Review*. https://doi.org/10.1016/j.childyouth.2018.04.030.

Womack, S. R., Taraban, L., Shaw, D. S., Wilson, M. N., & Dishion, T. J. (2018). Family turbulence and child internalizing and externalizing behaviors: Moderation of effects by race. *Child Development, 0*(00), 1–16. https://doi.org/10.1111/cdev.13103.

Zahn-Waxler, C., Shirtcliff, E. A., & Marceau, K. (2008). Disorders of childhood and adolescence: Gender and psychopathology. *Annual Review of Clinical Psychology, 4*, 275–303. https://doi.org/10.1146/annurev.clinpsy.3.022806.091358.

Ziol-Guest, K. M., & Dunifon, R. E. (2014). Complex living arrangements and child health: Examining family structure linkages with children's health outcomes. *Family Relations*. https://doi.org/10.1111/fare.12071.