



Concern over internal, external, and incidence validity in studies of childcare quantity and externalizing behavior problems

Journal:	<i>Child Development Perspectives</i>
Manuscript ID	2016-032.R2
Wiley - Manuscript type:	Article
Keywords:	Child care quantity, Externalizing problems, Validity
Abstract:	<p>Literature reviews have concluded that extensive time in early childcare is associated with high levels of externalizing behavior problems. We are concerned with three domains of validity in the work underlying these conclusions. Regarding internal validity, most studies rely on covariate-adjusted correlations, a suspect approach for ensuring unbiased results. We review studies on childcare quantity and externalizing problems using rigorous approaches to reduce selection bias. Results are mixed, and inconsistent with a causal hypothesis in studies using multiple statistical approaches. Regarding external validity, the field has relied too heavily on U.S. samples. We call for more international replications allowing for the study of socio-political variations. Regarding incidence validity, study designs have the widest relevance when structured to address the opportunities and constraints that make up the realities of family life, today. We point the field toward asking questions about childcare quantity that maximize validity in each of these three domains.</p>

**Concern over internal, external, and incidence validity in studies of childcare
quantity and externalizing behavior problems**

For Review Only

Abstract

Literature reviews have concluded that extensive time in early childcare is associated with high levels of externalizing behavior problems. We are concerned with three domains of validity in the work underlying these conclusions. Regarding internal validity, most studies rely on covariate-adjusted correlations, a suspect approach for ensuring unbiased results. We review studies on childcare quantity and externalizing problems using rigorous approaches to reduce selection bias. Results are mixed, and inconsistent with a causal hypothesis in studies using multiple statistical approaches. Regarding external validity, the field has relied too heavily on U.S. samples. We call for more international replications allowing for the study of socio-political variations. Regarding incidence validity, study designs have the widest relevance when structured to address the opportunities and constraints that make up the realities of family life, today. We point the field toward asking questions about childcare quantity that maximize validity in each of these three domains.

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 3

Not long after women's workforce participation started to rise in 1970s, developmental psychologists became embroiled in debate over potential harms of non-maternal childcare for children's behavioral development (e.g., 1-3). Three decades later, multiple literature reviews on the topic have been inspired. Aiming to provide a conceptual framing to this literature, Belsky (4) argued that "early, extensive, and continuous" non-maternal care heightens risk for the development of disobedient and aggressive behaviors, a conclusion echoed in a review by Jacob (5). Lamb and Ahnert (6) arrived at a similar, if more tentative, conclusion allowing for the caveat that quality may moderate quantity effects. More recently, Huston et al. (7) argued that findings are mixed, although for middle-class White, non-Hispanic, quantity effects on externalizing problems are real.

While these conclusions may lead parents to worry, as is indeed evident from the most cursory search of parenting forums on the Internet, the economic and social realities of family life in the 21st century have made non-maternal childcare an integral institution within thriving societies. Indeed, the gap between women and men's workforce participation continues to narrow in the U.S. and internationally (8). In turn, nearly two-thirds of children under the age of 5 are in regular non-maternal childcare in the U.S., and this proportion is considerably higher in most other wealthy nations (9). It is within this sociocultural reality of family life in the 21st century that we offer an alternative conclusion to those from recent reviews, arguing that too little attention to internal and external validity may be biasing much of the scientific and public thinking on the topic. To conclude, we critique the incidence validity of work on childcare quantity and offer our perspective on promising pursuits for the field.

How confident are we that childcare quantity "effects" are causal? ...not very

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 4

As evidence of a correlation between time in non-maternal childcare and behavior problems first emerged, developmental scientists (e.g., 2,3) drew attention to an internal validity concern: the selection problem. Childcare choices are influenced by family constraints, privileges, and preferences. Thus, when children are not randomly assigned to childcare conditions, we must be concerned that unobserved characteristics of children, families, and their wider ecologies bias correlation-based estimates of childcare effects. Even when children can be randomly assigned to childcare programs, ensuring that parents (in the control and treatment groups) will use no other forms of non-parental childcare is often infeasible.

In the empirical literature to date, the primary response to concern over selection bias has been expanding the range of measures included; researchers routinely use multivariate methods (e.g., ordinary least-squares regression) with large numbers of covariates, from children's temperament to parent personality to family socioeconomics. An impressive array of alternatives to the causal hypothesis have been measured and controlled. One strength of this approach is that associations may be estimated efficiently (i.e., small standard errors). There is, however, a trade-off: even with extensive covariate sets, covariate-adjusted associations are susceptible to bias. The trouble is that "alternative explanations are never completely enumerable" in non-experimental designs (10). And, causal estimates may be biased by including the wrong, or poorly measured, covariates (10). Consequently, there is reason to be skeptical that covariate-adjusted correlations provide unbiased estimates of the causal effects of childcare quantity.

Our argument, however, is not that researchers should avoid probing causality in non-experimental childcare studies. We argue quite the opposite. Indeed, several statistical methods beyond covariate adjustment are available for probing the causal

hypothesis with correlational data (11). With regression discontinuity, instrumental variable, and difference-in-difference techniques, researchers can exploit natural experiments in which factors beyond participants' control (e.g., policy changes) result in their assignment to "treatment" versus "control" conditions in a manner that closely approximates randomization. Within-person fixed-effect methods exploit changes over time in children lives, allowing researchers to rule out bias caused by between-child heterogeneity; this approach can be used to examine whether increases in childcare hours are associated with increases in externalizing problems. Similarly, sibling fixed-effects can be used to compare siblings with differing amounts of time in childcare, ruling out bias caused by between-family heterogeneity.

Although these methods rarely provide opportunities for causal inference that match a well designed randomized experiment, they are powerful tools for expunging plausible alternatives to the causal hypothesis. A limited number of childcare studies have, in fact, been conducted using these methods to test the hypothesis that time in care causes externalizing problems. Here, we briefly review these studies, focusing primarily on the question: are childcare quantity "effects" on externalizing problems robust or fragile once subjected to methods that conservatively control for selection bias? Even if childcare quantity results prove null or fragile – substantially varying in statistical significance, effect size, or direction of effect – one cannot logically accept the null hypothesis and conclude that childcare quantity does not harm children. Yet, suspension of belief in the causal hypothesis may be in order if results are not robust to correctly specified estimation strategies that offer internal validity strengths (12).

For our review, we used three literature search strategies. First, we examined all articles cited in the most recent review (7). Second, we searched PsycInfo. Third, we included relevant articles referenced in manuscripts, or referencing manuscripts,

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 6

retrieved using the first two strategies. Further details on this search strategy are described in a supplementary document available online. Among the articles that studied associations between childcare quantity (broadly defined as age of entry, hour per week, episodes per interval, childcare or preschool vs. parental child care, or any combination thereof), we included those using statistical methods more conservative than conventional covariate-adjustment approaches. Thirteen studies met this criterion (13-25).

In Table 1, we list the study authors, the country in which data were collected, the statistical method, the “treatment” of interest (i.e., center-based care, center-based and informal forms of child care, or increased access to center-based care due to policy shifts), the counterfactual (i.e., exclusive parental care, higher quantities of parental care, or a mixture of parental care and non-parental care other than center-based care), and a brief summary of results. Although we cannot fully detail sampling, context, and methodological variations across these studies, we note two key factors before turning to the results of interest.

First, in large part due to the country-differences, there are variations in: (a) average age of entry ranging from 3 months (21) to 14 months (25); and (b) average weekly hours (e.g., at 36 months, ranging from was 36.7 in one US study (21) to 29.45 in a Norwegian study (23)). Second, while many studies use nationally representative samples (13-19), one is predominantly low SES (20) and others are more concentrated with middle and high SES families (21-25). Even among those samples that are nationally representative, variations in “treatment” effects across SES or race/ethnicity are not addressed with other methods than multivariate regression. Thus, we do not address whether cause probing yields dissimilar findings across these groups. With these key factors in mind, when best practice methods for probing

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 7

causality with non-experimental data are employed, the evidence is quite mixed as to whether time in childcare causes externalizing behavior problems.

[Insert Table1 here]

In six of the studies, time in childcare predicted increased problems. On the other hand, six studies reported null results or, in one case, indicated that more time in childcare caused *fewer* problems. In addition, one study produced mixed results in which quantity of care was associated with more problems in some models and unrelated to behavior problems in other models.

Also revealing are the results of the seven studies that used both covariate-adjusted correlations and more conservative methods; in five of the seven, results differed – in significance or direction of effect – when moving from covariate adjustment to more rigorous methods (i.e., 18, 20, 21, 23, 25). In four of these five, more time in care was associated with more problems when using conventional covariate-adjusted methods, but was null or associated with fewer problems when using a more rigorous method. In the fifth study (21), the results were mixed when using methods beyond covariate-adjustment. These within-sample failures to replicate are reminders of the problem of relying exclusively on measured covariates to control for selection into childcare. It is possible, of course, that varying results across methods were due to differences in efficiency (i.e., size of standard errors). However, this appeared to be the case in only one study (25), and for only one of three tests of the causal hypothesis these authors conducted.

Nonetheless, as is the case with covariate adjustment, these methods are valuable only to the extent that their assumptions are met. The use of fixed-effects, for example, requires variability in childcare hours over time, or between siblings. With regard to instrumental variable estimation, finding valid instruments (i.e., the external

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 8

factors causing variations in childcare) in child development research is difficult. Moreover, fixed-effects and instrumental variable estimates are not as efficient (i.e., larger standard errors) as ordinary least-squares estimates.

In light of this, we agree with calls to replicate results across multiple methods (26). In turn, the field must consider whether the results of childcare quantity studies are robust or fragile to methodological variation? Our review suggests they are fragile, giving good reason to be skeptical that childcare quantity effects are causal. Nonetheless, we offer the caveat that we do not fully explore sampling or design variations across these studies. It is worth noting, for example, that four of the six studies reporting harmful consequences of time in childcare (13, 14, 16, 19) compared a “treatment” of publicly-funded center-based childcare or preschool with a mixed counterfactual of parental care and/or alternative non-parental care arrangements. And, four of the six come from a single province in Canada (13-16). These factors may be true moderators of the causal effects. More generally, we argue that replication as a scientific tool has been underutilized in the study of variability in childcare effects, particularly international replication.

How much do we know about childcare outside of the US? ...too little.

Replication has gained considerable attention in the psychological sciences as a tool for examining robustness and for scrutinizing the reproducibility of results (e.g., 26, 27), but less attention has been given to its value as a tool for exploring the conditions under which findings differ systematically. When researchers are interested in whether a study’s results can be successfully reproduced, the aim is to examine the robustness of findings under conditions that approximate the original study. However, sensitivity of results to conditions that differ across studies is also highly informative. Replication with variation in study conditions allows for the study

of “predictable exceptions” (28); thereby, replication can help us answer not only *if* extensive time in childcare causes behavior problems but also *when* and *where* it might be harmful, helpful, or inconsequential. Answers to these latter questions are of both scientific and policy relevance, and ultimately offer improved external validity for the cumulative knowledge on childcare quantity.

Childcare researchers have been concerned with the when and where questions. Association between time in care and externalizing problems may be moderated by factors such as childcare quality, child characteristics, and family contexts (7). Less attention has been given to whether this association varies across sociopolitical contexts, despite longstanding recognition of the potential value of international work on this topic, given variations in government support and regulation of care (e.g., 29). In the most recent review (7), 21 of the 29 articles cited are based on children in the U.S. Moreover, more than half--13--of the U.S. studies are based on one sample, the NICDH Study of Early Child Care and Youth Development. It is difficult to generalize these findings beyond the U.S., because the U.S. is a peculiar sociopolitical context among developed nations.

For decades, many advanced economies other than the U.S. have offered: job-guaranteed, paid parental leave; family allowances to offset costs of child rearing; universal health services; and social benefits that do not expire. Indeed, these were salient political factors highlighted over 20 years ago by childcare researchers (29). In addition, today, many countries have regulations on work hours, extensive holidays, and universal childcare beginning in infancy (23). Such policies affect how early, how extensive, and how continuous children’s childcare experiences are.

Presently, childcare quantity has been studied in Australia, Canada, Denmark, Japan, the Netherlands, Norway, Slovenia, Portugal, and the U.K. (7, 14, 15, 16, 23-

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 10

25, 30-32). Yet, none of these countries has received even a quarter of the empirical attention that the U.S. has. Given limited research beyond the U.S., we are currently restricted to consider comparisons of quantity effects across sociopolitical contexts as case studies. For example, it is tempting to speculate that risks associated with early childcare may not present in countries with comprehensive packages of family and child support policies such as in Norway, but arise only in less supportive contexts. Yet, some of the more rigorous evidence that childcare causes behavior problems comes from Canada, which also provides relatively progressive child- and family supports. Of course, countries often differ along multiple dimensions, including Canada and Norway, and as the international research base grows, the potential to isolate individual dimensions of economic and sociopolitical context improves. As an example, there is a vast literature based on international large-scale assessments of student achievement, allowing for both within-country and between-country moderators to be isolated (e.g., 33).

As the international evidence base grows, we offer suggestions for variables worth isolating. The childcare evidence from within countries empirically informs some these suggestions. For example, international variations in typical age of entry, average hours of care, and quality of care are worth attention. In terms of child and family characteristics, evidence from the U.S. on the SES and race/ethnicity as a moderator (7) is worth pursuing across and within international work. In addition, we should give careful attention to variations in work-family policy that affects children's lives such as health care, cost of childcare, distribution of public childcare (e.g., universal versus targeted), and social benefits.

Are we asking the right questions about time in childcare? ...sometimes.

It has been proposed that incidence validity should be a central criterion for evaluating the importance of developmental research; that is, for how large a portion of the population are the answers to our questions relevant (34)? Increasingly, the incidence validity of studies comparing young children in exclusive maternal (or even exclusive parental) care with those in other forms of childcare is low, because non-parental childcare is an integral part of life for most families in modern economies. Indeed, most recent studies go beyond comparisons of non-parental care versus exclusive parental care. This is not to say that the field has remained in lockstep with social trends in childcare use.

For employed mothers with young children, the average hours of employment is near 35 hours per week in the U.S. (35), and while the standard deviation for these mothers is close to 13 hours per week, the distribution is skewed; more than twice as many women work full-time as part-time (36). Nationally representative data on childcare indicate a similar distributional shape (37). Moreover, evidence suggests that, typically, within-child changes in hours are fairly small for children near the median (e.g., 38). An evidence base that is high in incidence validity requires methods that direct careful attention to these most highly represented childcare experiences in the population, in terms of quantity of care (and the typical level of changes). Yet, it is not uncommon for researchers to either (a) divide childcare quantity into broad categories that include relatively wide ranges of time in care (e.g., for part-time childcare versus full-time childcare) or (b) rely exclusively on linear specifications of time in care (for exceptions, see e.g., 21, 23, 38, 39). Both of these strategies may limit our ability to study the type of incremental differences that commonly occur for children within the highest incidence range. For example, while evidence on non-linearity is mixed, evident non-linear associations have been characterized by small

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 12

differences in the problem behaviors of children between about 25 and 35 hours of care but increasingly larger differences between about 35 and 50 hours (21, 23).

If our findings are to inform family decisions on the value of decreasing or increasing children's hours in care, we should give further attention to incremental differences for children within high incidence ranges. More generally, a valid research direction for the field is our continued pursuit of answers to how families and societies should best organize and support children's childcare experiences in ways that help children and families thrive. Questions of quantity remain relevant, including issues of how early in life children start childcare and how many hours they spend there, but only so far as our study designs structure these questions according to the opportunities and constraints that make up the realities of family life. Studies that inform families and policy decision makers on ways to enhance parents' chances for making optimal choices for their children's care deserve our greatest attention.

As one case in point, consider a meta-analysis of preschool programs by Schindler et al. (40). These authors found no behavioral differences between children attending preschool programs and control children, if preschools did not have specific components focused on fostering healthy behavioral development. Yet, when programs had integrated social skills training, they reduced children's behavior problems. Also worth highlighting are studies of public childcare scale-up in which researchers have taken advantage of natural quasi-experiments to study impacts on children's behavior (e.g., 13, 41). Neither of these areas of study is ideal for evaluating whether early, extensive, and continuous time outside in childcare leads to heightened problems, because parents in the control or comparison conditions generally have choices over alternative (non-maternal) care arrangements. Yet, the ecological and incidence validity of such studies is high as families make decisions

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 13

concerning where, at what age, and for how many hours their child will attend childcare.

What recommendations can we offer on promising pursuits for the field? ...use the right methods, in the right places, to answer the right questions.

Our perspective on the current state of the childcare research addressing time in care and its consequences for children's behavioral development is, ultimately, a call to reconsider how, where, and what questions we pursue as a field. Regarding the "how," we continue to give too little attention, when designing studies and interpreting results, to internal validity threats. If we consider only studies that have pushed control for selection bias beyond covariate-adjusted correlations, the evidence that childcare quantity negatively affects levels of externalizing behaviors is fairly weak. Regarding the "where," the current evidence is heavily U.S.-centric. We must prioritize international work on childcare, expanding opportunities to examine systematic variations across sociopolitical contexts, if only because research on childcare quantity has potential implications for both parents' choices and policy decisions globally.

Finally, regarding the "what," our study questions should be framed around modern economic and social realities. Today, the incidence validity of our studies of childcare will be maximized through analyses that allow us to examine incremental differences among those who spend considerable portions of their day in non-parental care. Questions of quantity remain relevant, including how early in life children should start childcare and how much time in childcare is best for promoting health development outside of their parents' care. Pursuing these questions with attention to internal, external, and incidence validity offers our best chance of informing policy and supporting families' efforts to promote their children's healthy development.

References

1. Belsky, J. (1986). Infant day care: A cause for concern? *Zero to three*, 6, 1-9.
2. Clarke-Stewart, K. A. (1989). Infant day care: Maligned or malignnant? *American Psychologist*, 44, 266-273.
3. McCartney, K., & Rosenthal, S. (1991). Maternal employment should be studied within social ecologies. *Journal of Marriage and the Family*, 53, 1103-1107.
doi: 10.2307/353014
4. Belsky, J. (2001). Emanuel Miller Lecture - Developmental risks (still) associated with early child care. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 42, 845-859. doi: 10.1111/1469-7610.00782
5. Jacob, J. I. (2009). The socio-emotional effects of non-maternal childcare on children in the USA: a critical review of recent studies. *Early Child Development and Care*, 179(5), 559-570. doi:10.1080/03004430701292988
6. Lamb, M. E., & Ahnert, L. (2006). Nonparental child care: Context, concepts, correlates, and consequences. In W. Damon, R. M. Lerner, K. A. Renninger, & I. E. Sigek (Eds.), *Handbook of child psychology: Volume 4: Child psychology in practice* (5th ed.). Hoboken, NJ: Wiley.
7. Huston, A. C., Bobbitt, K. C., & Bentley, A. (2015). Time spent in child care: How and why does it affect social development? *Developmental Psychology*, 51, 621-634. [doi:10.1037/a0038951](https://doi.org/10.1037/a0038951)
8. OECD. (2012). *Closing the Gender Gap – Act Now*. Paris: OECD.
9. OECD. (2014). *PF3.2: Enrolment in childcare and pre-schools*. Paris: OECD.

Retrieved from:

http://www.oecd.org/els/soc/PF3_2_Enrolment_in_childcare_and_preschools.pdf

10. Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.
11. Miller, P., Henry, D., & Votruba-Drzal, E. (2016). Strengthening Causal Inference in Developmental Research. *Child Development Perspectives*, 10, 275-280.
doi: 10.1111/cdep.12202
12. Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology*, 13, 219-240. doi: 10.1080/13501780600733376
13. Baker, M., Gruber, J., & Milligan, J. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, 116, 709-745.
doi:10.1086/591908
14. Baker, M., Gruber, J., & Milligan, K. (2015). *Non-cognitive deficits and young adult outcomes: The long-run impacts of a universal child care program* (No. w21571). National Bureau of Economic Research.
15. Kottelenberg, M. J., & Lehrer, S. F. (2013). New evidence on the impacts of access to and attending universal child-care in Canada. *Canadian Public Policy*, 39, 263-286.
16. Kottelenberg, M. J., & Lehrer, S. F. (2014). Do the perils of universal childcare depend on the child's age? *CESifo Economic Studies*, 60, 338-365.
17. Loeb, S., Bridges, M., Bassok, D., Fuller, B., & Rumberger, R. W. (2007). How much is too much? The influence of preschool centers on children's social and cognitive development. *Economics of Education Review*, 26, 52-66.
doi:10.1016/j.econedurev.2005.11.005
18. Jaffee, S. R., Van Hulle, C., & Rodgers, J. L. (2011). Effects of nonmaternal care in the first 3 years on children's academic skills and behavioral

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 16

functioning in childhood and early adolescence: a sibling comparison study. *Child Development*, 82(4), 1076-1091. doi:10.1111/j.1467-8624.2011.01611.x

19. Magnuson, K. A., Ruhm, C. J., & Waldfogel, J. (2007). Does prekindergarten improve school preparation and performance? *Economics of Education Review*, 26, 33-51. doi:10.1016/j.econedurev.2005.09.008
20. Crosby, D. A., Dowsett, C. J., Gennetian, L. A., & Huston, A. C. (2010). A tale of two methods: comparing regression and instrumental variables estimates of the effects of preschool child care type on the subsequent externalizing behavior of children in low-income families. *Developmental Psychology*, 46, 1030-1048. doi:10.1037/a0020384
21. McCartney, K., Burchinal, M., Clarke-Stewart, A., Bub, K. L., Owen, M. T., & Belsky, J. (2010). Testing a series of causal propositions relating time in child care to children's externalizing behavior. *Developmental Psychology*, 46, 1-17. doi: 10.1037/a0017886
22. Morrissey, T. W. (2009). Multiple child-care arrangements and young children's behavioral outcomes. *Child Development*, 80, 59-76. doi: 10.1111/j.1467-8624.2008.01246.x
23. Zachrisson, H. D., Dearing, E., Lekhal, R., & Toppelberg, C. O. (2013). Little evidence that time in child care causes externalizing problems during early childhood in Norway. *Child Development*, 84, 1152-1170. doi:10.1111/cdev.12040
24. Zachrisson, H. D., & Dearing, E. (2015). Family income dynamics, early childhood education and care, and early child behavior problems in Norway. *Child Development*, 86, 425-440. doi:10.1111/cdev.12306

25. Dearing, E., Zachrisson, H. D., & Nærde, A. (2015). Age of entry into early childhood education and care as a predictor of aggression: Faint and fading associations for young Norwegian children. *Psychological Science, 26*(10), 1595-1607. doi: 10.1177/0956797615595011
26. Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology, 50*(11), 2417-2425. doi:10.1037/a0037996
27. Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716. doi:10.1126/science.aac4716
28. Lindsay, R. M., & Ehrenberg, A. S. C. (1993). The design of replicated studies. *The American Statistician, 47*(3), 217-228. doi:10.1080/00031305.1993.10475983
29. Scarr, S., & Eisenberg, M. (1993). Child care research: Issues, perspectives, and results. *Annual Review of Psychology, 44*, 613-644. doi: 10.1146/annurev.ps.44.020193.003145
30. Coley, R. L., Lombardi, C. M., & Sims, J. (2014). Long-term implications of early education and care programs for Australian children. *Journal of Educational Psychology, 107*, 284-299. doi:10.1037/a0037456
31. Torres, N., Veríssimo, M., Santos, A. J., Monteiro, L., Figueiredo, M., & Vaughn, B. E. (2015). Quantity of group child care, behavior problems, and prosocial behaviors: A study with portuguese preschoolers. *Early Education and Development, 26*(8), 1145-1165. doi: 10.1080/10409289.2015.1024994
32. Vidmar, M. (2015). Contextualising ECEC research: the case of Slovenia. *International Journal of Early Years Education, 23*, 313-328. doi: 10.1080/09669760.2015.1078726

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 18

33. Seaton, M., Marsh, H. W., & Craven, R. G. (2009). Earning its place as a pan-human theory: Universality of the big-fish-little-pond effect across 41 culturally and economically diverse countries. *Journal of Educational Psychology, 101*, 403-419. doi: 10.1037/a0013838
34. Fabes, R. A., Martin, C. L., Hanish, L. D., & Updegraff, K. A. (2000). Criteria for evaluating the significance of developmental research in the twenty-first century: Force and counterforce. *Child Development, 71*(1), 212-221. doi:10.1111/1467-8624.00136
35. Blundell, R., Bozio, A., & Laroque, G. (2013). Extensive and intensive margins of labour supply: work and working hours in the US, the UK and France. *Fiscal Studies, 34*, 1-29. doi: 10.1111/j.1475-5890.2013.00175.x
36. Berik, G., & Kongar, E. (2012). Time Use of Mothers and Fathers in Hard Times: The US Recession of 2007-09. *Levy Economics Institute, Working Papers Series, 726*.
37. Flanagan, K., and West, J. (2004). Children Born in 2001: First Results From the Base Year of the Early Childhood Longitudinal Study, Birth Cohort (ECLS–B) (NCES 2005–036). U.S. Department of Education, Washington, DC: National Center for Education Statistics.
38. NICHD Early Child Care Research Network. (2003). Does amount of time spent in child care predict Socioemotional adjustment during the transition to kindergarten? *Child Development, 74*, 976–1005. doi:10.1111/1467-8624.00582
39. Solheim, E., Wichstrøm, L., Belsky, J., & Berg-Nielsen, T. S. (2013). Do time in child care and peer group exposure predict poor socioemotional adjustment in Norway? *Child Development, 84*, 1701-1715. doi:10.1111/cdev.12071

40. Schindler, H. S., Kholoptseva, J., Oh, S. S., Yoshikawa, H., Duncan, G. J., Magnuson, K. A., & Shonkoff, J. P. (2015). Maximizing the potential of early childhood education to prevent externalizing behavior problems: A meta-analysis. *Journal of School Psychology, 53*, 243-263.
[doi:10.1016/j.jsp.2015.04.001](https://doi.org/10.1016/j.jsp.2015.04.001)
41. Datta Gupta, N., & Simonsen, M. (2010). Non-cognitive child outcomes and universal high quality child care. *Journal of Public Economics, 94*, 30-43.
[doi:10.1016/j.jpubeco.2009.10.001](https://doi.org/10.1016/j.jpubeco.2009.10.001)

For Review Only

VALIDITY CONCERN IN STUDIES OF CHILD CARE QUANTITY 20

Table 1

Studies of time in care/use of child care, using quasi-experimental methods (accounting for unobserved selection into care) to estimate associations with externalizing problems

Authors	Country	Method	Treatment	Counterfactual	Results
Baker, et al. (13)	Canada	DID	Access to CC	MIX	Increased problems ^a
Baker, et al. (14)	Canada	DID	Access to CC	MIX	Increased problems
Kottelenberg & Lehrer (15)	Canada	IV	CC & INF	EPC	Increased problems ^a
Kottelenberg & Lehrer (16)	Canada	DID	Access to CC	MIX	Increased problems ^a
Loeb et al. (17)	US	IV	Early CC/INF Entry	Late CC/INF Entry	Increased Problems ^{b,d}
Jaffee et al. (18)	US	SFE	CC	LTC	NS ^{b,c}
Magnuson et al. (19)	US	IV	CC	MIX	Increased Problems ^c
Crosby et al. (20)	US & Canada	IV	CC & INF	EPC	Reduced Problems ^{b,c}
McCartney, et al. (21)	US	CFE	CC & INF	LTC	NS & Increased Problems ^{b,c}
Morrissey (22) ^e	US	CFE	CC & INF	LTC	NS
Zachrisson et al. (23)	Norway	CFE, SFE	CC	LTC	NS ^{b,c}
Zachrisson & Dearing (24)	Norway	CFE	Early CC Entry	Late CC Entry	NS
Dearing et al. (25)	Norway	IV	CC	LTC	NS ^{b,c}

Note. IV = instrumental variable; DID = difference-in-difference; CFE = child fixed effects; SFE = sibling fixed effects; CC = center care; INF = informal, non-parental care; EPC = exclusive parental care; LTC = less time in care. ^aThese authors estimate effects of ECEC only for children in two-parent families. ^bAuthors also estimated covariate-adjusted correlations. ^cResults differed in direction or significance when compared with covariate-adjusted correlations. ^dThese authors report concerns with the validity of their instrument. ^eHours in care is a covariate in this article.

Online supplement**Concern over internal, external, and incidence validity in studies of childcare quantity and externalizing behavior problems****Eric Dearing & Henrik Daae Zachrisson**

First, we conducted a search in PsycInfo for peer-review scholarly articles written in English since 1970 (excluding qualitative and interview-studies), using the search terms ("child care" OR "non-parental care" OR "non-parental child care" OR daycare OR preschool OR childcare OR "early education") AND (externalizing OR "behavior problems" OR aggression OR hyperactivity OR non-compliance) in the abstract. This search retrieved 1452 articles. We manually searched these abstracts for descriptions of analyses/discussions of child care quantity, defined as either age of entry or dosage of child care, hours per week or equivalent, or use vs. not use of child care., or combinations of these. This resulted in 40 selected studies, including the three reviews we cited in the manuscript. We then compared this list with studies cited in Huston et al., (2015), finding that our search retrieved 22 additional empirical articles on quantity, however with 4 of them published in 2015 and thus could not be included by Huston et al. We then read the method sections of the articles retrieved from our search as well as those cited in Huston et al., searching for analytical approaches going beyond covariate adjusted regression analyses. This resulted in 10 articles. Finally, we searched articles citing the 10 articles included, and retrieved additional three articles from economics. Including these, we eventually arrived at the list of 13 studies initially presented in Table 1.