Does Childcare Make Children More Stressed? A Meta-Analysis of the Effects of Childcare on Children's Cortisol Levels

Erik Eliassen

Oslo and Akershus University College of Applied Sciences

August 1, 2012

Abstract

Prior research has indicated that being in childcare causes children to be stressed. However, research on the effect of childcare on children's stress levels, as measured by the hormone cortisol, has yielded inconsistent results. While some studies have found a strong positive association between childcare and increased cortisol levels, other studies have found no such association, or a negative association.

The aim of this study was to review the literature and to perform two meta-analyses on the effect childcare has on children's cortisol levels. A total of 10 studies (N=738) were reviewed and provided the basis for the analyses. I used two designs for the meta-analyses: the first design estimated an effect on a relative scale, and was based on a comparison of the *difference* between cortisol levels in the morning and afternoon for children at home compared to children in childcare. The second design estimated an effect on an absolute scale, and was based on comparing the *average* of the morning and afternoon cortisol levels of children at home compared to children in childcare.

The findings of this study suggest that, on a relative scale, the cortisol levels of children in childcare did not decrease as much throughout the day as they did for children at home ($d = 0.31$, 95% CI [0.20, 0.41]). However, on an absolute scale, children in childcare did not display significantly higher cortisol levels on average than children at home ($d = 0.06$, 95% CI $[-0.15, 0.26]$). Overall, these findings suggest that the effect of childcare on children's cortisol levels is to some extent present, but the effect size is arguably weak.

Acknowledgements

Several people have played important parts in the process of performing this study. First of all, I would like to thank my advisor Lars Gulbrandsen for his keen insights and vital support to the project. In a field where quantitative research is not always associated with superlatives, Lars greeted the project with enthusiasm and played an important role throughout the process. Also, I would like to thank my skilled co-advisor Ellen Os, who unfortunately – due to no fault of her own – I did not get to work with as much as I would have liked. Both Ellen and Lars agreed to continue working on the project to completion even after it was in danger of getting postponed, which is a reflection of their commitment and dedication to the field.

I am especially indebted to Karina Corbett, who opened my eyes to the wonderful world of statistics and quantitative research. She also helped nurture me back to health during a difficult year and played a pivotal role in the completion of the study. Without her I would not have been able to finish on time, and her contributions to this project have been invaluable.

I would also like to thank my good friend Erik Hynne for taking the time out of a busy schedule to proofread the final draft of this paper and for providing last-minute suggestions.

Lastly, I would like to express my gratitude to my parents for supporting me through a long education which has allowed me to focus most of my attention on my studies.

Contents

1 Introduction

Ever since the Norwegian government in 2005 announced its aim of providing childcare for all children under compulsory school age (Office of the Prime Minister 2005), the country has seen an expansive development within the kindergarten sector (Winsvold & Gulbrandsen 2009). Since 2004, the percentage of 1- to 2-year-olds in center-based childcare has nearly doubled, from 42% (Sæther 2010) to 80% in 2011 (Statistics Norway 2012c; Statistics Norway 2012d). In total, about 90% of all children living in Norway, between 1 and 5 years of age, are currently enrolled in either childcare centers run privately or by municipality (Statistics Norway 2012a). This development was further strengthened by the Norwegian government's passing of a law that gives all children under compulsory school age, who have turned at least one within August of the year they apply, a statutory right to kindergarten admission (Norwegian Ministry of Education and Research 2008).

As childcare is increasingly becoming the norm in our society, concern has also been expressed about how this relatively new way of organizing childhood is affecting the children. One of the most important aspects of early childhood education research is arguably that of children's health and well-being in relation to childcare, but in Norway there is a research gap on the issue (Gulbrandsen, Johansson & Nilsen 2002; Borg, Kristiansen & Backe-Hansen 2008). Particularly children's stress levels in these environments have been a focus of attention, as research from other countries has indicated that children might be more stressed in childcare than they are at home (eg., Legendre 2003; Watamura, Donzella, Alwin & Gunnar 2003; Dettling, Gunnar & Donzella 1999). Such findings have partially been responsible for sparking an ongoing public debate in Norway about whether childcare is harmful for children. Although the media coverage has contributed to a increased public awareness of potential health issues related to childcare, the debate has often times been lacking in nuance (eg., Tveitereid 2008; Melgård 2008; Sørensen 2011; Olsø 2011). In spite of this, research findings have been mixed, with some studies finding large effects in stress related to childcare, while other studies have found smaller effects, no effect or opposite effects. The aim of this study is therefore to review the literature and to synthesize the study results in meta-analysis, to try to assess whether childcare really does make children more stressed, and if so, to determine the degree. In the first part of the study I will discuss theoretical underpinnings for why one might expect childcare to be stressing, followed by a description of the biological causes of stress and how these relate to health outcomes. The second part of the study will consist of a review of the included studies, followed by a description of the methods used in this study. Lastly, the results will be presented and discussed accordingly.

1.1 Factors in Childcare That May Have a Negative Impact on Children

There are many positive aspects to childcare, both socially and developmentally. For instance, childcare centers can provide rich and stimulating environments for play, learning and social interactions. In Norway, kindergartens are bound by the Kindergarten Act (Norwegian Ministry of Education and Research 2005) to provide "children under compulsory school age with good opportunities for development and activity in close understanding and collaboration with the children's homes" (p. 1) while demands to quality, content and learning is regulated by a national framework plan (Norwegian Ministry of Education and Research 2006). Childcare centers can therefore be particularly important for children coming from disadvantaged families, and moderate negative effects related to challenging home conditions (Melhuish 2004). However, there are also factors in childcare that could cause stress and carry more of a negative impact. The effects of these may be short-term, but could also potentially cause more long-term problems. Important risk factors related to the physical and social environment, as well as emotional factors are therefore discussed in the following section.

1.1.1 Physical environment

Noise Noise is part of everyday life, and most of the time we effectively filter it out. But for most people who have either worked in childcare or have children in childcare it is no surprise that these environments can be highly noisy. Although some noise is generally harmless, loud noise or chronic exposure to noise can be draining and exhaustive (Evans 2006). According to research by the Norwegian Labor Inspection Authority (2005), as much as 75% of all employees working in Norwegian kindergartens report having work environments that have high noise levels. Given that children mostly occupy the same areas as the adults who work there, it is likely that the children are exposed to the same noise levels – or perhaps even higher noise levels than the caregivers. Studies have also shown that chronic noise exposure is associated with increased stress levels in children (for review, see Evans 2006).

Cramped spaces Another aspect of the physical childcare environment that might differ from the home environment is the physical space. Although childcare centers might be larger than the average home, the available space is also shared by many individuals, both children and adults. In the traditional kindergarten model the childcare center is divided into different 'classrooms' that the children may largely be confined to. These classrooms are often relatively small considering the number of children inhabiting them and are likely significantly more cramped than the children's home environments. Recent trends within the Norwegian kindergarten sector have shown development and organization moving away from the traditional kindergartens, and towards so-called open-plan kindergartens. These are kindergartens without the traditional 'classrooms,' opting instead for larger and more centralized spaces that can accommodate larger groups of children (Vassenden, Thygesen, Brosvik, Alvestad & Abrahamsen 2011). According to Vassenden et al. (2011), roughly half of all kindergartens in Norway today are open-plan kindergartens, while the majority of kindergartens developed after 2005 have been organized as open-plan kindergartens. Although they found no clear association between kindergarten-model and size of kindergarten – in terms of number of children attending – it is nevertheless the case that most of the largest kindergartens (in terms of both size and population) are organized as either open-plan, or as partially open-plan (Vassenden et al. 2011). This tendency towards larger groups and open landscapes reduces children's possibilities to "hide away" when needed, and might be more taxing on the children's abilities to cope with their social environment. This was also shown by Rappolt-Schlichtmann et al. (2009) who found that going from large group environments to smaller group contexts reduced the children's over-all stress levels.

1.1.2 Social environment

Lack of tailoring to individual needs Many childcare centers are rigidly organized in their day-to-day routines, offering children little flexibility in deviating from planned activities. In many respects, children in childcare therefore have limited control over many aspects of their environments and are in varying degrees subject to changes that could make them more stressed. For instance, childcare centers may have nap schedules that do not necessarily coincide with the children's sleeping patterns, have fixed daily routines that might interfere with the children's own activities or interrupt their play, or situate them in settings they might not feel comfortable in. Children's stress responses in relation to such conditions will likely vary, as certain children may be less adaptable to childcare environments due to their specific temperaments or characteristics. Dettling, Parker, Lane, Sebanc and Gunnar (2000) studied the relationship between child characteristics and stress levels and found that emotional negativity or limited self-control predicted cortisol elevations, while Watamura, Donzella, Alwin, and Gunnar (2003) found a similar association connected to social fear. A lack of tailoring to these individual needs thus seems to be a likely source of stress for many children.

Child-to-staff ratios Several issues seem to arise from the way in which childcare is structured. These factors are known as structural processes and include child-to-staff ratios and group sizes, among other things. These are considered to be closely related to quality of care (Norwegian Ministry of Education and Research 2008-2009) and studies have shown that low child-to-staff ratios generally tend to be associated with less distressful behavior in children (Vandell & Wolfe 2000). Interestingly, Legendre (2003) found the opposite to also be true in that having many caregivers was also associated with increased stress levels.

High turnover and staff instability Since children generally seem to benefit from having a sufficient number of caregivers available (Norwegian Ministry of Education and Research 2008-2009; Vandell & Wolfe 2000) they are therefore also dependent on the staff being stably present. One of the main factors behind staff instability is high turnover. In the four-year period between 2003 and 2007, the turnover rate for Norwegian kindergarten teachers was one out of six (16.2%), while the turnover rate for uneducated kindergarten staff was over one out of three (37.7%) (Gulbrandsen 2009). Furthermore, sickness absence also contributes heavily to staff instability and is currently at 11% within the kindergarten sector (Statistics Norway 2012b). Sickness absence can also have the side effect of resulting in replacement staff that may lack necessary education or pedagogical training – or even cause understaffing. The resulting instability could be unsettling for the children – perhaps especially for those who lose their closest caregivers. Insecure children, or very young children may also be extra vulnerable.

1.1.3 Emotional factors

According to Bowlby (1977), children are first and foremost attached to their primary caregiver(s), generally the mother. As a result of this symbiotic bond, any unwilling separation from this primary caregiver could cause severe anxiety and emotional distress. Although this bond gradually wanes over time as we become increasingly independent, it is of vital importance in early life. As mentioned earlier, many children in Norway attend kindergarten from a very young age. In 2011, 2599 infants (children under 1) were enrolled in childcare (Statistics Norway 2012c). Children between the age 0 and 24 months constitute nearly 17% of the kindergarten population in Norway (Statistics Norway 2012), while the children between zero and 36 months constitute 36% of all children in childcare. Consequently, a large quantity of children attend childcare before they have acquired language or learnt how to walk. The severity of being separated from one's primary caregivers at such early ages will depend on a range of factors and is likely to affect children differently. Research has shown that children's reactions to these daily separations unfold in complex ways, and that the severity of children's responses is dependent on moderating factors such as quality of secondary care, duration of 'goodbye' rituals, etc. (Klein, Kraft & Shohet 2010)

1.2 Stress in Children

In order to understand if childcare is making children more stressed, it is also necessary to understand what stress is, how it is defined and conceptualized, and ultimately how it is likely to relate to children. Cohen, Kessler and Gordon (1995) describe the psychological perspective as oriented around a person's "perception and evaluation of the potential harm posed by stimuli (stressors or events)" (p. 121). These perceptions are thought to arise from situations where "imposed demands exceed the felt ability to cope with these demands" (p. 121). From a biological perspective, stress is defined as "a state of threatened homeostasis, which is reestablished by a complex repertoire of physiologic and behavioral adaptive responses" (Chrousos 1998, p. 311), where homeostasis is defined as the precondition to life through the maintenance of a "complex dynamic equilibrium" (p. 311).

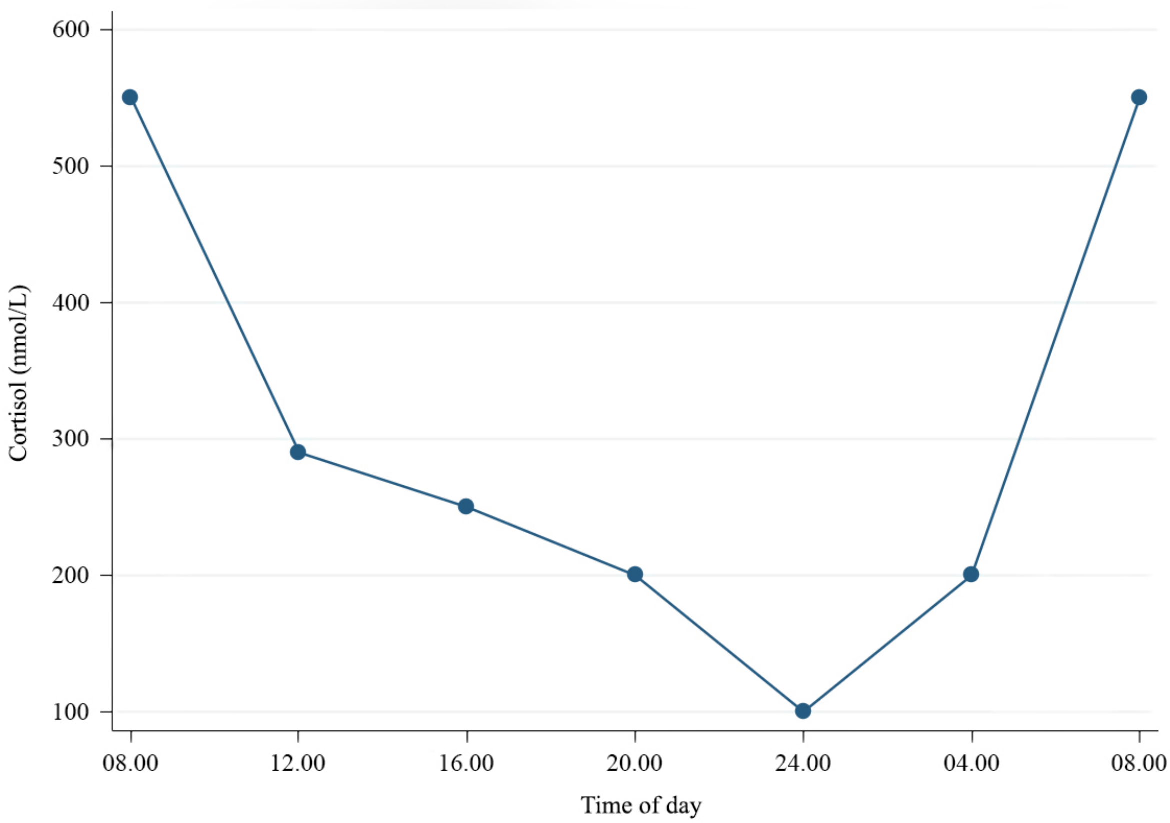
One significant difference between these two definitions is that the former is based on subjective experiences while the latter is oriented around more objective measures. The psychological perspective therefore seems to presuppose a cognitive awareness of one's current state and an ability to evaluate one's capacities. However, the extent to which young children are capable of such evaluative processes or the degree to which they are consciously aware of their resourcefulness is not clear. This therefore raises questions with regards to how children experience stress in line with the psychological definition.

These two definitions also pose methodological problems for how to accurately measure stress, especially when it comes to young children. Generally, perceived stress has been measured with the Perceived Stress Scale [PSS] (Cohen, Kamarck & Mermelstein 1983), which relies on questions and responses. However, since many children are too young to either express themselves accurately, or may be unable to understand stress conceptually, researchers have been looking for other ways of inquiry. The biological perspective, on the other hand, lends itself more readily to stress measurement in children since its definition hinges upon physical responses. In this regard, researchers have identified that when we become stressed our bodies increase production of the hormone cortisol, which allows stress to be measured more objectively (Kirschbaum & Hellhammer 1994), and will therefore be addressed in the following section.

1.3 Cortisol and the Circadian Rhythm

Since children's stress-responses are not always externalized or easily observable, researchers have instead attempted to measure children's stress levels through biological markers, namely by measuring the hormone cortisol (eg., Watamura, Coe, Laudenslager & Robertson 2010; Legendre 2003; Dettling, Gunnar & Donzella 1999) which has been identified as a reliable stress marker (Kirschbaum & Hellhammer 1994). Cortisol is a steroid hormone that is produced by the hypothalamic-pituitary-adrenocortical [HPA] axis, which mediates stress-sensitive processes in the body (Rodriguez, Mayo & Sainz 2000). The hormone affects physiological changes in most of the main organ system (Adam & Kumari 2009) and has a production that normally follows a specific cycle or pattern known as the circadian rhythm or diurnal pattern (Watamura 2004). The circadian rhythm is considered to be a fundamental biological phenomenon and has a cycle of 24 hours (Hardeland 2000). Typically, cortisol levels peak shortly after wake-up, decrease throughout the day, and reach base levels in the evening (Watamura 2004). It has been widely studied (Rodriguez, Mayo & Sainz 2000) and is not influenced by short-term manipulations of the day-night cycle, or by feeding-schedule (Krieger as cited in Rodriguez et al. 2000). According to de Weerth, Zijl and Buitelaar (2003), children are born without a cortisol circadian rhythm, but acquire it during their first year of life. However, they emphasize that studies have not consistently agreed on its time of appearance in infants, with estimates ranging from two weeks to nine months.

Figure 1 – Example of expected diurnal cortisol pattern



The graph was redrawn and approximated from Wilson et al. (1998), and is set to represent a normal diurnal cortisol pattern.

1.4 Health Effects Associated With Elevated Cortisol Levels

Large-scale epidemiological research is increasingly relying on salivary cortisol measurement to assess public health due to the important function of the HPA axis in relation to physiological changes relevant to health (Adam & Kumari 2009). Our bodies are dependent on short-term activations of the HPA, but frequent, extreme or chronic HPA activation is correlated with negative health outcomes (Adam & Kumari 2009). In other words, normal bodily functioning is dependent on cortisol levels remaining within an optimal range (Herbert et al. 2006). Chronic conditions in cortisol elevations can cause a loss of the diurnal cortisol rhythm and may in either case contribute to disturbance in brain function (Herbert 2006). Healthy HPA axis functioning appears to depend on strong circadian rhythms, meaning that the cortisol production follows the normal diurnal cortisol pattern (Adam & Kumari 2009), while flattened or reversed diurnal cortisol patterns have shown to be associated with chronic and acute psychosocial stress (for review, see Adam & Kumari

2009). Longitudinal studies have also shown that frequent activation of the HPA axis is associated with metabolic syndrome (Brunner et al. 2002), and Halligan, Herbert, Goodyer and Murray (2004) found that exposure to postnatal depression predicted elevated cortisol in adolescent offspring, therefore suggesting that adverse experiences in early age might alter later hormone levels. As is clear from research, stress can have serious health consequences. However, it is also important to keep in mind that not all stress is negative, if contained in a controllable state, stress can also result in experiences of being in control, and even give rise to feelings of pleasure and reward (Chrousos 1998).

1.5 Meta-Analysis

Several studies have attempted to answer the question of whether childcare makes children more stressed, but the results have not been consistent. In such cases, meta-analyses are used to "synthesize the available evidence for a given question to inform policy." (Borenstein et al. 2009, p. xxv). In other words, performing a meta-analysis is an attempt to make sense of the big picture, to gather any available information on a given topic to inform us of what works and what doesn't. As Deeks et al. (2011) note, an important aspect of performing a meta-analysis is that it has the potential of settling conflicting claims since it does not merely rely on a single study's findings, but carries the weight of a multitude of studies. A meta-analysis is typically a two-stage process (Deeks et al. 2011). The first stage consists of calculating a summary statistic, also known as an effect size, for each of the selected studies. The effect sizes describe the observed effects from the individual studies. The second stage consists of calculating a summary effect estimate based on a weighted average of the individual effects.

More formally, a meta-analysis has the advantage of increasing power, and improving precision (Deeks, Higgins & Altman 2011). Statistical power is the probability of not making a type II error, the error of not rejecting the null hypothesis when it is actually false (Dietz & Kalof 2009). Since sample size is closely related to power, pooling the sample sizes from the individual studies in a meta-analysis will therefore naturally increase the probability of finding an effect when there really is one. In terms of this study, performing a meta-analysis will hopefully increase the likelihood of determining whether or not children really are more stressed in childcare than they are at home.

1.5.1 Prior meta-analyses

There have been earlier meta-analyses studying the effects of childcare on children's stress levels, notably by Vermeer and IJzendoorn (2006) and Geoffroy, Côté, Parent & Séguin (2006). Although similar in approach, there are differences between the current study and the prior meta-analyses. An important difference is that both of the prior meta-analyses mainly studied changes in diurnal cortisol patterns, but diurnal cortisol patterns are only one aspect of a complex stress regulatory system. This study will provide a complementary analysis that will also compare average cortisol levels on an absolute scale. These designs will be further addressed in the methods chapter. Additionally, the current study will provide updated analyses by including the multiple new studies that have been published since 2006. It will be the largest meta-analysis to date, with a total sample size of 738 children compared to the Vermeer and IJzendoorn study (N=303) and the Geoffroy et al. study (N=541). Lastly, it is worth noting that the Vermeer and IJzendoorn study appears to suffer from some miscalculations in the analysis, as the authors have estimated non-significant study results (n=2) to be equivalent to absolute zero effect (r=0.00). This misconception inadvertently affects the effect size incorrectly, and partly undermines the rigorousness of their analysis.

2 Methods

2.1 Literature Searches

A search of the relevant terms was performed in the databases PsycINFO, Medline, Pubmed and ERIC. Search terms consisted of "day care/daycare/day-care," "child care/childcare/child-care" combined with "stress," "cortisol" and "hypercortisone." When the databases provided the option, the search terms were exploded, enabling the databases to return results not just from the specified search terms, but also from similar or related search terms. Finally, various lists of references from relevant studies were also searched. Ideally, one would attempt to locate the unpublished studies as well to counter-act the problem of publication bias (Borenstein et al. 2009) – which will be addressed in the discussion chapter – but finding and retrieving such studies is very difficult. An alternative strategy to avoid this problem is to exclude studies that have not been through a peer-review process. The rationale behind such an approach is that non-peer-reviewed (and consequently unpublished) studies are considered to be of lower quality. The latter approach was therefore implemented in this study. Although advocated by some (eg., Weisz et al. 1995), others (eg., Borenstein et al. 2009) have contested the legitimacy of such an approach.

2.1.1 Inclusion criteria

After the relevant literature had been identified through the literature search, the studies were examined and included in the meta-analyses if they met the following inclusion criteria:

1. Cortisol levels must have been measured both in childcare and at home, as in a cohort study or a randomized control trial.
2. Time points have been matched - cortisol must have been measured at approximately the same time of day in both childcare and at home.
3. Cortisol measurements must have been collected at multiple time points, both in childcare and at home (only required for the first meta-analysis)
4. The study must have been published in an international, peer-reviewed journal.
5. Sufficient data must have been provided for effect sizes to be calculated.
6. The samples have been drawn from a population of children attending center-based childcare

2.2 Literature Review

As an integral part of any meta-analysis, the process of appraising and reviewing of the studies should also be carried out in a systematic manner to minimize bias (Glasziou, Irwig, Bain & Colditz 2001). For the current literature review, the studies were assessed by the following questions:

1. Which population was studied?
2. Was the quality of care determined?
3. How and when did cortisol sampling occur?
4. Were steps taken to minimize bias?
5. Were there any methodological limitations?
6. What were the main findings?

2.3 Meta-Analysis

2.3.1 Design

Two designs were selected to address the research question of whether children are more stressed in childcare than they are at home. The first design compared the diurnal patterns across the childcare group and the home group. Specifically, the morning cortisol values were subtracted from the afternoon values, and the difference was compared between the groups. This design was adapted from what is known as 'difference-in-difference' (Meyer 1995; What Works Clearinghouse 2008), and provided a way of comparing the cortisol fluctuations in childcare and at home on a relative scale. However, there is an important difference between the adaptation and the original. In a 'difference-in-difference' design pre-exposure measurements are subtracted from post-exposure measurements, but in this case the morning measurements (equivalent to "pre-measurements") from childcare are not strictly speaking 'pre-exposure' since the children have already been exposed to childcare for a short while before being measured. Nevertheless, the chosen design gives good indication on how children's diurnal patterns differ from childcare to home; if the difference between morning and afternoon levels is lower for children in childcare than children at home, one would interpret this as meaning that children in childcare were more stressed. This interpretation is irrespective of initial starting values. Thus, even if children in childcare, for instance, start off with much lower cortisol levels, as long as their afternoon levels have not decreased more than for children at home on an absolute scale, one would conclude that they were more stressed. Lastly, it is also important to note that due to this specific design any additional cortisol samples, apart from morning and afternoon samples, were disregarded from this analysis.

Although diurnal cortisol patterns play an important role in the stress regulatory system, known as the HPA axis (Adam & Kumari 2009), the patterns alone do not provide information on how high the cortisol levels are on average. The second meta-analysis therefore compared the average cortisol levels in childcare with the average cortisol levels at home, without taking into account whether the levels changed from morning to afternoon. This meta-analysis therefore included all available cortisol measurements due to the simplicity of the design. The procedures of both the chosen designs are explained mathematically in detail under the 'Mathematical Formulas' section.

2.3.2 Fixed effect and random effects models

There are two main statistical models that are frequently used in meta-analysis, both with different assumptions with regards to the interpretation of the between-study variation in effect sizes. The random effects model assumes that the observed variation in effects is explained by actual differences in true effects. The fixed effect model, however, assumes that the observed between-study variation is due to some kind of measurement error, and that the underlying effect - given an infinitely large sample – would be the same for all studies (Borenstein 2009). Choosing the appropriate model also has implications for dealing with the potential of bias considering that the random effects model is more affected by funnel plot asymmetry since small studies with low precision are given relatively more weight than in the fixed effect model (Duval 2005). Taking into account that these models may produce different results, the meta-analyses will be performed and displayed with both underlying models for comparison. The summary effect based on the fixed effect model is signified graphically in the meta-analyses as "I-V Overall", while the summary effect based on the random effects model is signified as "D+L Overall". The appropriateness of the fixed effect model will be tested by a chi-square (χ^2) test. Should the χ^2 test prove non-significant above the alpha ($\alpha = 0.05$) level, the summary effect based on the fixed effect model will be reported. Conversely, should the χ^2 test prove significant below the α=0.05 level – indicating that the dispersion in effect sizes stems from actual heterogeneity – the summary effect based on the random effects model will be reported, since the random effects model allows for heterogeneity by assuming that the underlying effects follow a normal distribution (Deeks et al. 2011). Should there be little difference between the outcomes of the two models, the interpretation of either will be the same.

2.3.3 Effect sizes

The individual measures within a meta-analysis are commonly referred to as effect sizes. Borenstein et al. (2009) define an effect size as "(...) a value that reflects the magnitude of the treatment effect or (more generally) the strength of a relationship between two variables" (p. 3). Effect sizes can be calculated in a number of ways, depending on the measure of choice. In accordance with Borenstein et al. (2009) the standardized mean difference (measured in Cohen's d) was deemed the most appropriate measure for these meta-analyses since the summary data collected from the primary studies were mainly means and standard deviations from two groups. Typically, the collected studies reported cortisol measurement data in nmol/L and µg/dL, in addition to either standard deviations [SD] or standard errors of the mean [SEM]. Furthermore, the standardized mean difference is a summary statistic that is frequently used when the selected studies all measure the same thing, albeit in different ways. In such cases it is necessary to standardize the results in order to make them directly comparable for the meta-analyses (Deeks, Higgins & Altman 2006). Specifically, the standardized mean difference provides a way of comparing the overlap and/or difference between the distributions of two comparison groups (Borenstein et al. 2009).

2.3.4 Mathematical formulas

In order for studies to be synthesized in meta-analyses they must provide sufficient data for effect sizes to be calculated or converted to same measure. Researchers often report findings differently, depending on research design, statistical methods of use, or purely by preference. Conversions will therefore have to be made, which makes it harder for the reader to go back and compare the measures used in the meta-analyses with the ones that are reported in the original studies. For the sake of transparency, I therefore provide the formulas I have used to make these conversions and calculations so that any effects I report may be double-checked methodically.

When cortisol measurements were taken at only one time point, and mean values and standard deviations were provided, effect sizes were calculated by:

$$d = \frac{xchildcare - xhome}{\sigma pooled} \tag{1}$$

where:

$$\sigma pooled = \sqrt{\frac{(nchildcare - 1)SDchildcare^2 + (nhome - 1)SDhome^2}{nchildcare + nhome - 2}}$$

(Geoffroy, Côté, Parent & Séguin 2006).

A modified version of the formula above was applied to the meta-analysis based on comparing average cortisol levels (Figure 6), where all time points were collapsed and mean values were compared:

$$d = \frac{\bar{x}childcare - \bar{x}home}{\sigma pooled} \tag{2}$$

where:

$$\bar{x}childcare - \bar{x}home = \left(\frac{1}{n}\sum_{i=1}^{n} xchildcare_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n} xhome_i\right)$$

and:

$$\sigma pooled = \sqrt{\frac{(nchildcare - 1)SD_w childcare^2 + (nhome - 1)SD_w home^2}{nchildcare + nhome - 2}}$$

Since the standard deviations cannot be correctly averaged simply by calculating their mean, the average standard deviation of the sample was instead calculated by squaring the weighted variances:

$$SD_w = \sqrt{\sum_{k=1}^{n} \frac{S_i^2(n_i - 1)}{n_i - 1}}$$

In the meta-analysis based on change in diurnal cortisol pattern, the morning levels were subtracted from the afternoon levels, and the difference in daycare was compared with the difference at home:

$$d = \frac{x\Delta childcare - x\Delta home}{\sigma_{post} pooled} \tag{3}$$

where:

$$x\Delta childcare - x\Delta home = (xchildcare_{post} - xchildcare_{pre}) - (xhome_{post} - xhome_{pre})$$

and:

$$\sigma_{post} pooled = \sqrt{\frac{(nchildcare - 1)SDchildcare^2_{post} + (nhome - 1)SDhome^2_{post}}{nchildcare + nhome - 2}}$$

(What Works Clearinghouse 2008)

Pre-suffixes indicate cortisol values measured in the morning, while post-suffixes indicate cortisol values measured in the afternoon.

As suggested by Geoffroy et al. (2006), I have used the following formula to convert reported t-tests into the effect measure of Cohen's d (standardized mean difference):

$$d = \frac{2t}{\sqrt{df}} \tag{4}$$

The variance of d is to good approximation calculated by:

$$V_d = \frac{n_1 + n_2}{n1n2} + \frac{d^2}{2(n1 + n2)} \tag{5}$$

and the standard error [SE] of d:

$$SE_d = \sqrt{V_d} \tag{6}$$

(Borenstein 2009)

In studies where standard errors of the mean [SEM] were reported instead of standard deviations [SD], standard deviations were calculated by:

$$SD = SEM\sqrt{n} \tag{7}$$

2.3.5 Assessing bias

It is crucial to be aware of the fact that several factors could bias a sample and ultimately affect the precision and reliability of the results (Borenstein et al. 2009). Bias assessment is therefore an important step in maintaining the integrity of the analyses. I have followed the recommendations of the Cochrane Bias Methods Group (Sterne, Egger & Moher 2011) advocating that formal bias tests should generally only be performed when there are at least 10 studies present in the meta-analysis. Fewer studies would cause the power of the tests to be too low to distinguish chance occurrences from asymmetry resulting from bias. In this study there were consequently 14 (N=738) and 13 (N=665) independent samples included in the conducted meta-analyses. The use of formal testing, such as by Egger's regression test and Duval and Tweedie's trim and fill method, were therefore deemed appropriate to help assess and quantify bias.

The funnel plot Although the funnel plot was originally proposed as a means of detecting publication bias, it has later been pointed out that the reporting of exaggerated effect sizes from small studies with low quality and low power could provide an alternative explanation of plot asymmetry (Sterne & Harbord 2004). Since the precision of the effect size estimates increases proportionally with the sample sizes (n), the funnel plot can be visually assessed in determining whether the sample is biased (Sterne & Harbord 2004): "in the absence of bias, results from small studies will scatter widely at the bottom of the graph, with the spread narrowing among larger studies" (p. 127). Asymmetrically scattered studies around the mean, following the shape of the plot with most of the studies cropping up at the top and missing towards the bottom, could indicate bias (Borenstein, Hedges, Higgins & Rothstein 2009).

Egger's regression test Considering that funnel plot asymmetry could also be a product of mere chance, a formalized test of bias was proposed by Egger. Egger's regression test for small-study effects tests the hypothesis that there is no linear association between the effect and its standard error (Sterne & Harbord 2004). If such a relationship does exist, the amount of bias is quantified (Borenstein 2005).

Duval and Tweedie's trim-and-fill A major strength of Duval and Tweedie's trim-and-fill approach is that it addresses the question of what the best estimate of the unbiased effect size is (Borenstein et al. 2009). The method is based on an assumption that an unbiased study selection will result in funnel plot symmetry, which is why the trim-and-fill method re-estimates the effect size by eliminating studies that have no symmetric counterpart (Duval 2005). Duval (2005) explains the process as follows:

> [W]e trim off the 'asymmetric' right-hand side of a funnel plot (...), assumed to be affected by publication bias, after estimating how many studies are in this outlying part; [we] use the symmetric remainder to estimate the true 'centre' of the funnel; and then replace the trimmed studies and their missing 'counterparts' around the centre. The final estimate of the true overall effect, and also its variance, are then based on the 'filled' funnel plot. (p. 128)

Although the method does attempt to correct for bias and produce a new effect size estimate, the adjustment should be regarded as a form of sensitivity analysis to assess the potential impact of missing studies rather than be interpreted as a definitive correction (Duval 2005).

3 Results

3.1 Literature review

The literature review yielded 22 studies in total. Ten of these met all the inclusion criteria, whereas twelve did not. Three of the included studies (Watamura et al. 2003, Oullet-Morin et al. 2010 & Vermeer et al. 2010) contained analyses on two independently sampled groups that were differentiated by age. Since these groups were independent of each other and therefore represented different populations, they were treated as separate 'studies' in the meta-analyses, even though they originated from the same research papers.

Table 1 shows descriptive statistics of all the study samples that were included in the meta-analyses. The table shows that an overweight of the samples were from the US (n=6), two were from Canada, while the rest were from European countries (n=6). The samples varied greatly in sample size, ranging from 16 children in the smallest sample to 132 children in the largest sample. The combined sample represented all the childcare age groups (extending what is normal in Norway), and included infants as young as two months old to children nearly the age of 9.

Table 1 – Table of study samples that were included in the meta-analyses

	Study	**Year**	**Age (months)**	**Size (n)**	**Country**
1.	Dettling et al.	1999	39-106	51	Switzerland
2.	Dettling et al.	2000	35-66	40	USA
3.	Quas et al.	2002	53-74	50	USA
4.	Legendre[*]	2003	18-40	73	France / Hungary
5.	Watamura et al.	2003	2-16	16	USA
6.	Watamura et al.	2003	16-38	28	USA
7.	Watamura et al.	2009	32-64	55	USA
8.	Groeneveld et al.	2010	20-40	45	The Netherlands
9.	Oullet-Morin et al.[y]	2010	36	107	Canada
10.	Oullet-Morin et al.[y]	2010	24	132	Canada
11.	Vermeer et al.[f]	2010	16-41	41	Basque Country
12.	Vermeer et al.[h]	2010	16-41	19	Basque Country
13.	Vermeer et al.	2010	24-41	25	The Netherlands
14.	Watamura et al.	2010	36-60	56	USA
				$N^2 = 665$	
				$N^1 = 738$	

[*] = Only included in meta-analysis based on change in diurnal cortisol pattern.
[y] = Only based on first day measurements in meta-analysis on diurnal cortisol pattern.
N^2 = Total sample size for meta-analysis based on means.
N^1 = Total sample size for meta-analysis based on change in diurnal cortisol pattern.
[f] / [h] = Differentiated between full-day childcare and half-day childcare within the same age group.

The studies that were not included in the meta-analyses, along with their exclusion reasons, are listed in Table 2 on the following page.

Table 2 – Table of study samples that were excluded from the meta-analyses

	Study	**Year**
1.	Lundberg et al.[a]	1993
2.	de Haan et al.[ax]	1998
3.	Tout et al.[b]	1998
4.	Watamura et al.[a]	2001
5.	Legendre[*]	2003
6.	Sims et al.[b]	2006
7.	Watamura et al.[a]	2004
8.	Li et al.[b]	2008
9.	Rappolt-Schlichtmann[b]	2009
10.	Gunnar et al.[c]	2010
11.	Sumner et al.[a]	2010
12.	Gunnar et al.[a]	2011

[a] = Effect size could not be calculated.
[ax]= Home references were only sampled once. Later childcare measurements were not matched.
[b]= No home reference was provided.
[c]= Sampled from home-based daycare population.
[*]= Only excluded from meta-analysis based on comparing average cortisol levels.

Five out of twelve (42%) of the excluded studies did not meet the inclusion criteria in terms of robustness of research design. These studies either collected cortisol samples in childcare but not at home (Rappolt-Schlichtmann 2009; Li et al. 2008; Sims et al. 2006; Tout et al. 1998), or provided insufficient home measurements (de Haan et al. 1998). Such methodological short-comings cause inferential problems since without home-references it can only be speculated that any observed cortisol elevations in childcare have also been caused by childcare, and not by other uncontrolled factors. The second large group (42%) of excluded studies were the ones that did not provide adequate information for effect sizes to be calculated (Gunnar et al. 2011; Sumner et al. 2010; Watamura et al. 2004; Watamura et al. 2001; Lundberg et al. 1993). This was partly true for the Legendre (2003) study as well, which provided sufficient information to be included in the meta-analysis based on diurnal cortisol patterns, but not for the meta-analysis based on average cortisol levels. Lastly, one study was excluded because the sample population was non-center-based (Gunnar et al. 2010).

Tables 3 through 12 on the following pages contain descriptions and reviews of the studies that were included in the meta-analyses. The section thereafter will include a summary of important aspects found in the review process.

Table 3 – Review of the Dettling et al. (1999) study

Study	Dettling et al. (1999)
Populations studied	The study sample was drawn from a population of children attending two urban preschool daycare centers, and one daycare center for school-aged children. 70 children were sampled, and were aged between 39 and 106 months.
Quality of care	The daycare centers were considered to be of high quality, receiving high scores on standard assessment quality rating scales. The authors did not specify the specific scales of use.
Cortisol sampling	Saliva cortisol samples were measured across two days, at home and in full-day childcare at mid-morning and mid-afternoon.
Minimized bias	
Methodological limitations	The participating children were not randomized into the study, but were selected by convenience. The study also suffers from high refusal and dropout rates. 58% of the asked parents from the preschool center gave permission, where 86% out of the children with permission ended up being sampled. At the school age center, 43% allowed sampling, of which 90% of these children ended up being present during sampling. Reasons for declining or being absent during sampling were not specified in the study. The children were also given a small amount of sugar sweetened Kool-Aid to help salivation, which could potentially cause bias. The methodological issues related to such practice will be discussed later.
Main findings	The study found that three and four year old children in daycare were more likely to have higher cortisol levels at mid-morning than older children, implying that younger children and children with more limited social skills were vulnerable to cortisol elevations. The study also replicated earlier findings, showing that children in childcare displayed reversed diurnal cortisol patterns, while children at home followed the expected circadian rhythm.

Table 4 – Review of the Dettling et al. (2000) study

Study	Dettling et al. (2000)
Populations studied	The children were sampled from a University based childcare center and included 21 children aged between 35-63 months (median age = 52). The comparison group did not attend childcare and consisted of 19 children aged 39-66 (median age=52). The study also sampled a home-based daycare group, but this group was discarded from the meta-analyses due to the inclusion/exclusion criteria.
Quality of care	The childcare center was considered to be a model for high quality care, bt the authors did not specify the specific rating scales of use.
Cortisol sampling	Saliva sampling occurred at mid-morning and mid-afternoon. Researchers undertook cortisol collection in childcare over two days.
Minimized bias	The children were not allowed to eat within 30 minutes prior to cortisol sampling, nor were they allowed to drink dairy products within 15 minutes of sampling.
Methodological limitations	62% of the comparison group ended up providing cortisol data. The study suffered from low sample sizes, with a childcare group of 21 children, and a comparison group (home group) of 19 children. The children were given a few grains of sugar sweetened Kool-Aid to stimulate salivation which could theoretically influence the results.
Main findings	The study found that children who received higher quality care displayed no differences in terms of cortisol secretion in childcare compared to home. Children who received lower quality care in daycare displayed a reversal of the diurnal cortisol pattern from morning till afternoon, compared to the expected decrease in cortisol levels across the day at home. The study also found child characteristics such as temperament to be predictors of higher cortisol secretion, especially for children considered to be more emotionally negative or exhibiting less self-control.

Table 5 – Review of the Quas et al. (2002) study

Study	Quas et al. (2002)
Populations studied	The study sample consisted of 50 children, aged 53-74 months (mean age = 61 months).
Quality of care	The quality of care was not rated, though given the families' middle-class backgrounds the quality of care was estimated to be above average.
Cortisol sampling	Salivary cortisol samples were collected one week prior to starting in kindergarten and one week after. The pre-kindergarten measurements were used as controls.
Minimized bias	Instructions were given to collect saliva samples following wake-up, before breakfast. The children were not allowed to eat or drink at least 15 minutes prior to sampling, as it could influence the results. Children with chronic medical conditions, and children on medications were excluded from the study. The study tested for differences between the children who were successfully sampled and the children who for varying reasons ended up being excluded from the analysis. The study looked at family backgrounds, ethnicity, education, and income levels to investigate the generalizability of the convenience sample. The sample was considered ethnically diverse and the families were generally well-educated and middle class.
Methodological limitations	A convenience sample was used and artificial sweeteners were permitted during sampling. The children were allowed some hard candy to make salivation easier.
Main findings	The study found that children with little preschool experiences prior to attending kindergarten, and children experiencing greater change in their routines transitioning from preschool to kindergarten, displayed the highest increases in cortisol.

Table 6 – Review of the Legendre et al. (2002) study

Study	Legendre et al. (2002)
Populations studied	113 children were sampled, aged 18-40 months (mean age = 30 months). Eight toddler groups were examined, six French and two Hungarian. The sample was drawn from full-day public childcare centers.
Quality of care	Quality was assessed by how well the childcare centers met standards of structural criteria like child-to-staff ratio, available space for play etc. The selected centers varied in these respects.
Cortisol sampling	The children were sampled across four days, spread over two weeks. Saliva sampling took place after wake-up at home, and before and after free play. Saliva was also collected during weekends for control measurements.
Minimized bias	The samples were scanned for extreme values and outliers were discarded from analysis. The main cause for the outliers was considered to be illness or medication. Saliva sampling was conducted during winter or spring to ensure that the children were familiar with the child group, personnel, facilities and daily routines.
Methodological limitations	Children were allowed some orange flavored sugar to help with salivation.
Main findings	The study found that while cortisol levels followed the expected diurnal cortisol pattern at home, with levels decreasing throughout the day, cortisol levels in childcare remained flat. However, these effects were varying across the eight selected daycare centers, indicating a relationship between cortisol levels and environmental factors. It was also found that cortisol elevations were associated with large group sizes, age differences, small play areas and – surprisingly – many caregivers.
Note	Due to limited saliva collection at home, 73 of the children from this study were included in meta-analysis.

Table 7 – Review of the Watamura et al. (2003) study

Study	Watamura et al. (2003)
Populations studied	The participating children were 20 infants aged 2-16 months (mean age = 10.4 months), and 35 toddlers aged 16-38 months (mean age = 28.8 months). The children attended full-day daycare.
Quality of care	The quality of care was not specified.
Cortisol sampling	The study examined salivary cortisol samples. The children were sampled either after their morning or afternoon naps, if napping occurred, but before feeding time. Toddlers provided their morning samples after a time of free play.
Minimized bias	Children who were ill, or receiving any medication likely to have an effect on cortisol were excluded. The study also controlled for circumstances related to possible interference caused by sugar sweetened Kool-Aid.
Methodological limitations	Out of 111 potential participants, parents of 67 children agreed to participate. 45% of the participating children were not present during sampling days. It is also not clear how the children were allocated to the study, as no randomization procedure for the sample selection was specified. The children were given a few grains of sugar sweetened Kool-Aid to stimulate their saliva flows.
Main findings	35% of the infants and 71% of the toddlers displayed a rise in cortisol levels during daycare, compared to 71% of the infants and 64% of the toddlers who displayed decreased cortisol levels at home. They also found that toddlers who interacted in more peer play exhibited lower cortisol, and that social fear was a predictor of higher cortisol secretion throughout the day. The study indicated that cortisol elevations were age-dependent as toddlers in general displayed higher cortisol levels than the infants. The effect did, however, decline with the years, indicating that cortisol elevations did not appear permanent since they gradually regressed.
Note	Due to inconsistent participation across the sampling days, the number of children (from this study) that were included in meta-analysis averaged 44 (16 infants and 28 toddlers)

Table 8 – Review of the Watamura et al. (2009) study

Study	Watamura et al. (2009)
Populations studied	The study sample consisted of 65 children between the ages of 36-72 months, attending full day childcare.
Quality of care	Childcare was rated as high by use of the Modified-Observational Rating of the Caregiving Environment scale.
Cortisol sampling	Cortisol was sampled mid-morning, mid-afternoon, and evening both in childcare and at home.
Minimized bias	The study tested for differences between the children who refused sampling and the ones who allowed sampling. The two groups did not differ in age or gender. Children who were ill, were on medication, or had recently been distressed were excluded from analysis. Extreme outliers were also discarded from analysis. No Kool-Aid or artificial sweeteners were permitted during saliva sampling.
Methodological limitations	The child populations were selected by convenience at not by randomization.
Main findings	The study replicated earlier findings of cortisol rising across the day in childcare while decreasing across the day at home, albeit in a lesser degree than earlier studies had reported. Quality of care and surroundings were found to be predictors of these elevations.
Note	Due to inconsistent participation across the sampling days, the number of children (from this study) that were included in meta-analysis averaged 55.

Table 9 – Review of the Groeneveld et al. (2010) study

Study	Groeneveld et al. (2010)
Populations studied	The population sample consisted of 45 children aged between 20-40 months (mean age = 32 months).
Quality of care	The quality of care was assessed by global quality rating scale ECERS-R. The selected centers varied in quality across a range.
Cortisol sampling	The children were sampled across four time points both in childcare and at home.
Minimized bias	The sample was randomly selected from a national sample of 250 childcare centers, of which 26 centers were willing to participate.The children were not allowed to eat or drink within 30 minutes of saliva collection, and were not given Kool-Aid or artificial sweeteners to help salivate.
Methodological limitations	Despite being a well-designed study, it suffers from the low participation rate. The main reasons for declining were reported to be uneasiness with being video recorded, disliking the notion of saliva sampling, or weariness of being labeled as a low quality center.
Main findings	Although the study did find quality to be a predictor or cortisol levels, the children nevertheless exhibited higher cortisol in childcare than they did at home, regardless of the quality of care they were given.

Table 10 – Review of the Oullet-Morin et al. (2010) study

Study	Oullet-Morin et al. (2010)
Populations studied	The study sampled two populations consisting of 155 two year olds (mean age = 2.28) and 116 three year olds (mean age = 3.25). The participants were drawn from a large population of 5337 families, of which 1027 were eligible for inclusion.
Quality of care	The quality of care was rated with the Infants and Toddlers Environment Rating Scale and the Early Childhood Environment Rating Scale (ECERS). The quality of care was considered to range between adequate to good.
Cortisol sampling	Saliva was collected morning and afternoon, across two days. Saliva was collected by swabbing the children's cheeks with a sorbette.
Minimized bias	The study attempted to control for confounders related to sleep issues that could affect cortisol levels. Bias was minimized by not allowing children to eat or drink within 30 minutes of saliva collection.
Methodological limitations	
Main findings	The study found that the two year olds displayed a flattening of their diurnal cortisol patterns compared to home, while the three year olds showed decreasing patterns both at home and in childcare. The study found childcare experience to be a predictor of rise in cortisol within the three year old group, since the more experienced children showed lower levels in childcare than they did at home. Conversely, inexperienced children displayed higher cortisol levels in childcare than they did at home.
Note	Inconsistent participation in sampling resulted in the number of children (from this study) to average 132 in the two-year old group and 107 in the three-year group. Only first day measurements were included in the meta-analysis based on diurnal changes due to the specific design.

Table 11 – Review of the Vermeer et al. (2010) study

Study	Vermeer et al. (2010)
Populations studied	The study included children drawn from two populations, one sample from Basque Country (n=60) and one sample from the Netherlands (n=25). The children were aged between 16-41 months (mean age = 32).
Quality of care	The quality of care was measured with the Early Childhood Environment Rating Scale-Revised Edition (ECERS-R). Eleven of the included centers were rated as being of low quality, thirty-four centers were rated as moderate, while only one center was rated as high quality.
Cortisol sampling	Cortisol was measured at mid-morning and mid-afternoon over two days.
Minimized bias	To reduce the possibility of biasing the results, the children were not allowed to eat or drink within 30 minutes of saliva collection. No Kool-Aid or artificial sweeteners were used to help salivation. The child groups were randomly selected, with one group from each of the 46 included childcare centers.
Methodological limitations	
Main findings	The study found that children did not display higher cortisol levels in childcare compared to home, overall. The authors speculate that factors other than structural quality and quantity of care are responsible for changes in children's diurnal patterns.

Table 12 – Review of the Watamura et al. (2010) study

Study	Watamura et al. (2010)
Populations studied	The study sampled 65 three-to-five year olds (mean age = 4.3 years).
Quality of care	Quality was assessed with the Early Childhood Environment Rating Scale-Revised (ECERS-R) and was rated to be very high, ranging from good to excellent..
Cortisol sampling	Saliva sampling occurred mid-morning, mid-afternoon and evening over two days.
Minimized bias	The children were not allowed to eat or drink within 30 minutes of saliva collection. Sampling took place at least one hour before lunch to avoid cortisol production onset by anticipation, and not within 30 minutes after napping. No Kool-Aid or artificial sweeteners were used in relation to saliva collection.
Methodological limitations	
Main findings	The study looked at the association between the cortisol elevations observed in childcare and health issues. The results indicated that the flattened cortisol patterns seen during childcare predicted lower anti-body levels and increased risk of illness.

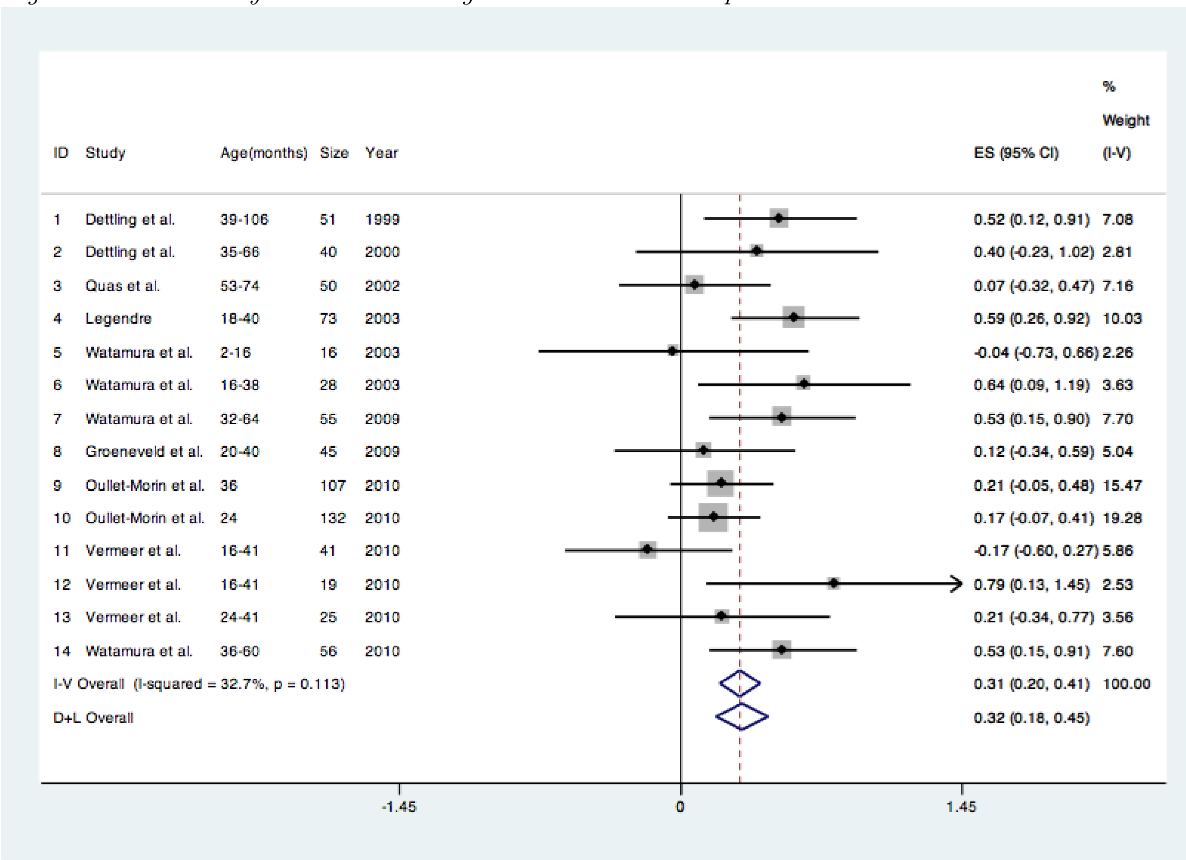
3.1.1 Summary

The quality of care provided by the childcare centers varied across the entire range. Five out of ten studies sampled children from high quality childcare centers (Watamura et al. 2010; Oullet-Morin et al. 2010; Watamura et al. 2009; Dettling et al. 2000; Dettling et al. 1999). Three studies reported the quality as varying (Vermeer et al. 2010; Groeneveld et al. 2010; Legendre 2003), while two studies did not assess childcare quality (Watamura et al. 2003; Quas et al. 2002). All the studies employed saliva sampling as a way of acquiring cortisol measurements, where half of the studies allowed Kool-Aid or artificial sweeteners to help salivation (Watamura et al. 2003; Legendre 2003; Quas et al. 2002; Dettling et al. 2000; Dettling et al. 1999). Four studies did not employ such practice (Watamura et al. 2010; Vermeer et al. 2010; Groeneveld et al. 2010; Watamura et al. 2009), while one study makes no mention of it (Oullet-Morin et al. 2010). The studies that allowed children sugar sweetened Kool-Aid (or similar) during sampling were also the oldest. This may be a result of a raised awareness around the methodological issues related to sweeteners, which will be further addressed in the discussion chapter. All the studies found childcare to be associated with changed diurnal cortisol patterns, with an exception of two subsamples (Watamura et al. 2003; Vermeer et al. 2010). Findings related to diurnal cortisol patterns were therefore mostly consistent, but the effects varied in strength.

3.2 Meta-Analysis Based on Change in Diurnal Cortisol Patterns

The results of the first meta-analysis – based on changes in children's diurnal cortisol patterns – are presented in Figure 2.

Figure 2 – Meta-analysis based on changes in diurnal cortisol patterns



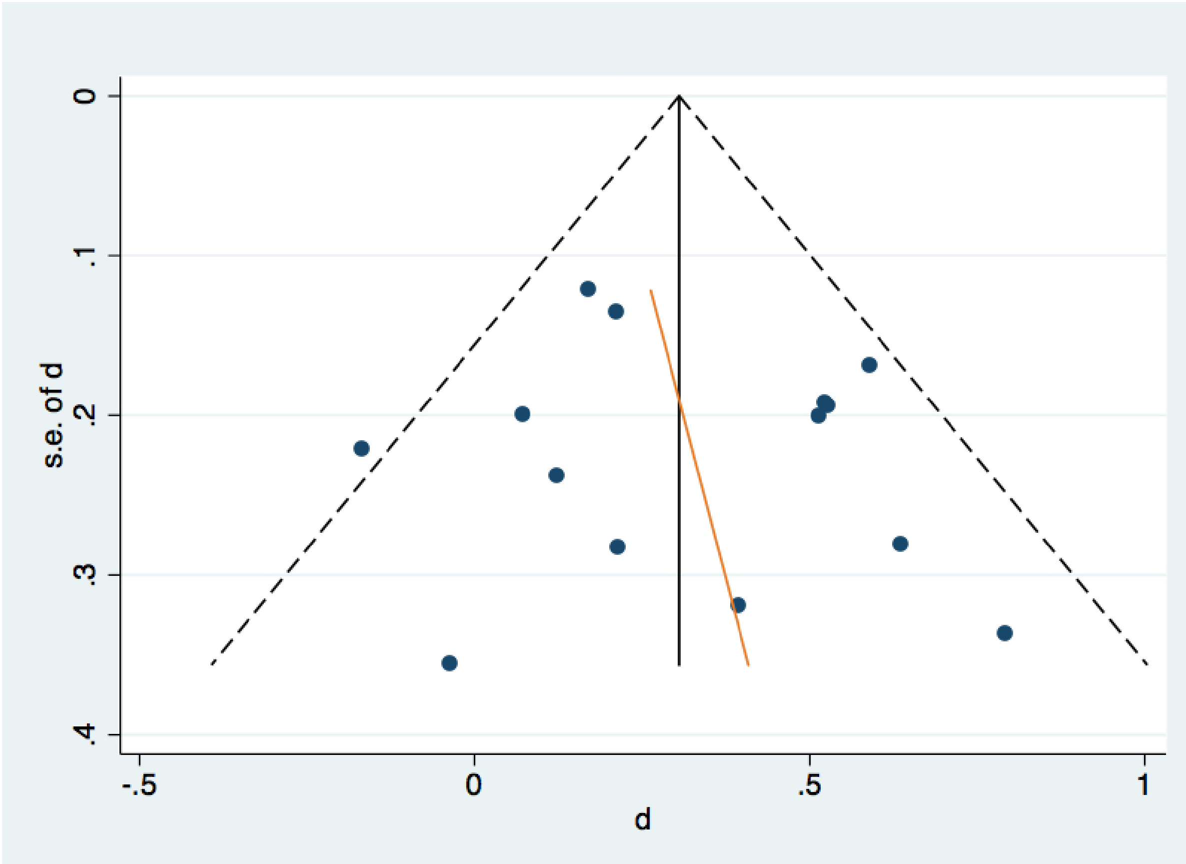
The meta-analysis is based on mid-morning and mid-afternoon cortisol values. N = 738. All effect sizes are in Cohen's d.

Fourteen samples, constituting a combined sample size of 738 children, produced an overall summary effect size of $d = 0.31$ (95% CI $[0.20, 0.41]$). The effect was significant ($z = 5.71, p < 0.001$), and does indicate that childcare attendance is to some extent associated with an increased risk of flattening or reversal of the diurnal cortisol patterns compared to home. Finally, a chi-square test was performed to test for heterogeneity, but the result was not significant ($\chi^2 = 19.33, df = 13, p = 0.113$).

3.2.1 Assessment of Bias

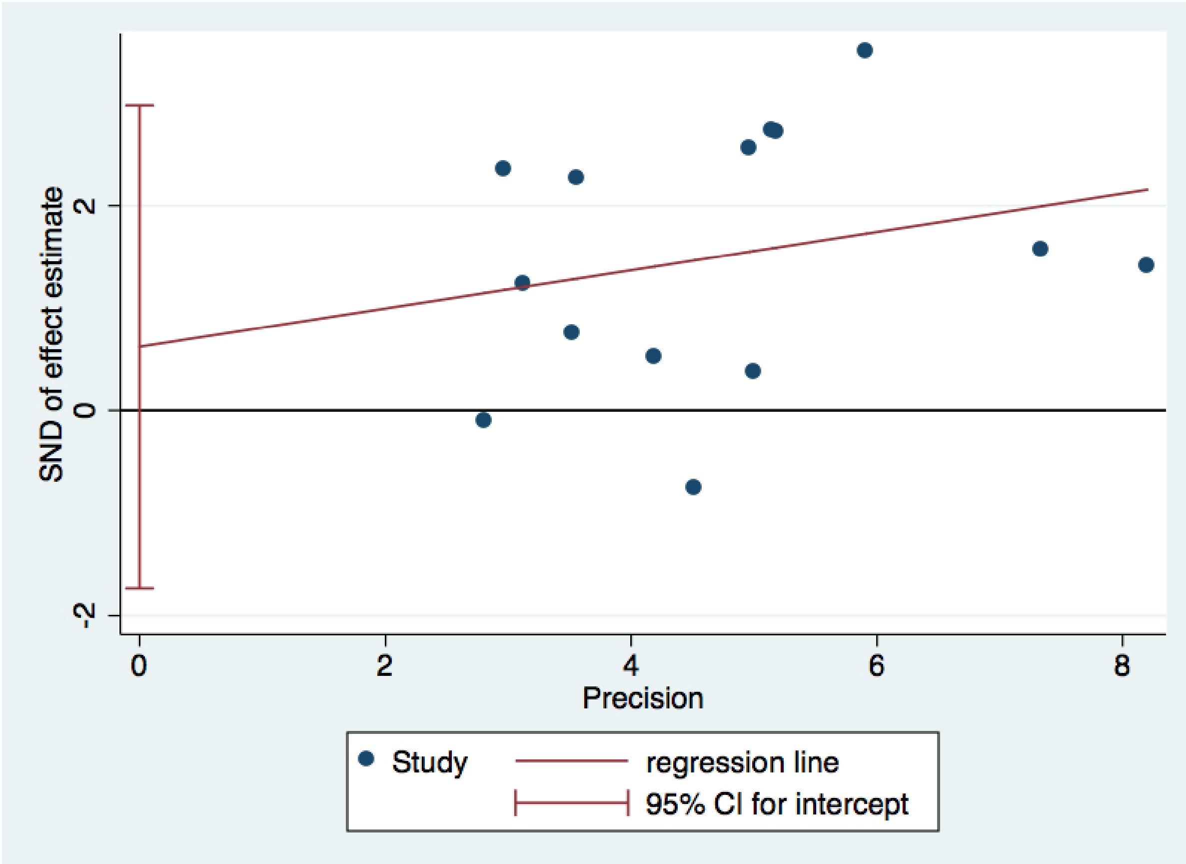
Figure 3 shows the funnel plot of the included samples from the meta-analysis. The funnel plot was fitted with a regression line to help visualize bias.

Figure 3 – Funnel plot with fitted regression line



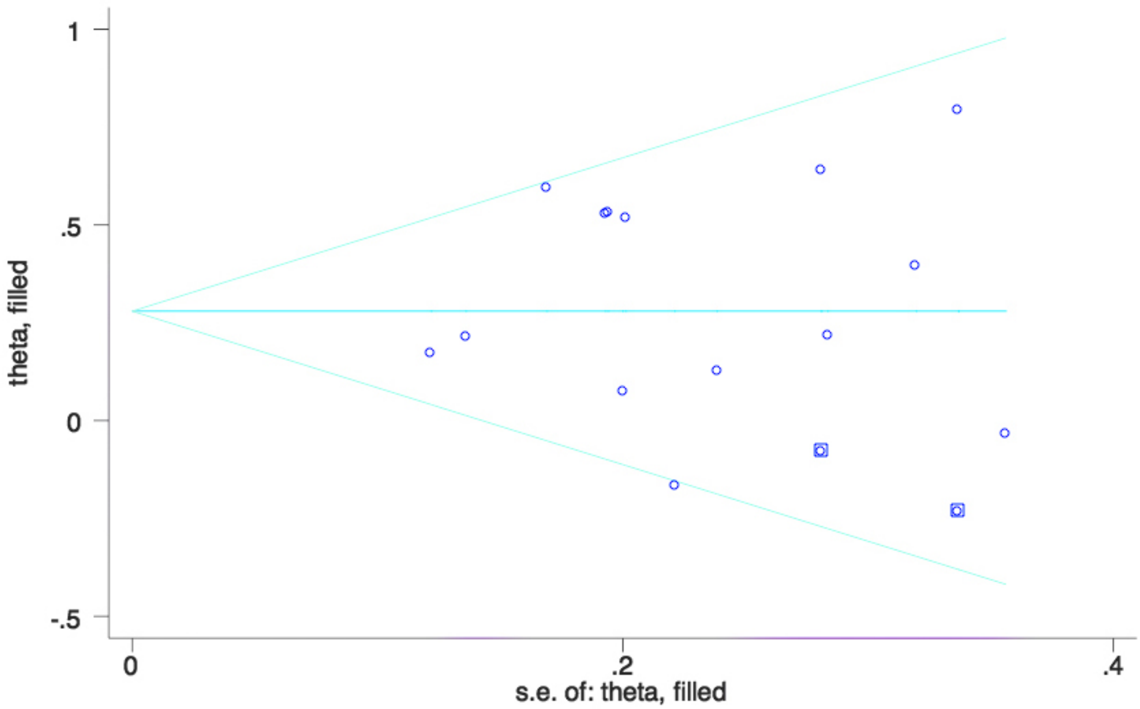
As we can see from Figure 3, the scatter does not completely follow the pattern we might expect if there was no bias present (as described in the method chapter). This impression is confirmed by the slight tilting of the regression line. Although the funnel plot hints towards some bias, it is not easy to determine the amount of bias just by visual assessment. An Egger's regression test for small study effects was therefore performed to quantify the bias:

Figure 4 – Egger's regression test for small-study effects



If there were no bias present we would expect the graphical presentation of the Egger's regression test to display a flat regression line. As is clear from both the funnel plot and the regression test for small study effects, the regression line is somewhat skewed. The slope was calculated to $\beta = 0.19$, 95% CI [-0.29, 0.66]. However, despite the regression line being somewhat skewed, we see from the test results that the confidence interval includes the null, indicating that the result is not statistically significant. The null hypothesis – the hypothesis that there is no bias present is therefore kept. This impression is further supported by the trim and fill method which is shown graphically in Figure 5.

Figure 5 – Filled funnel plot

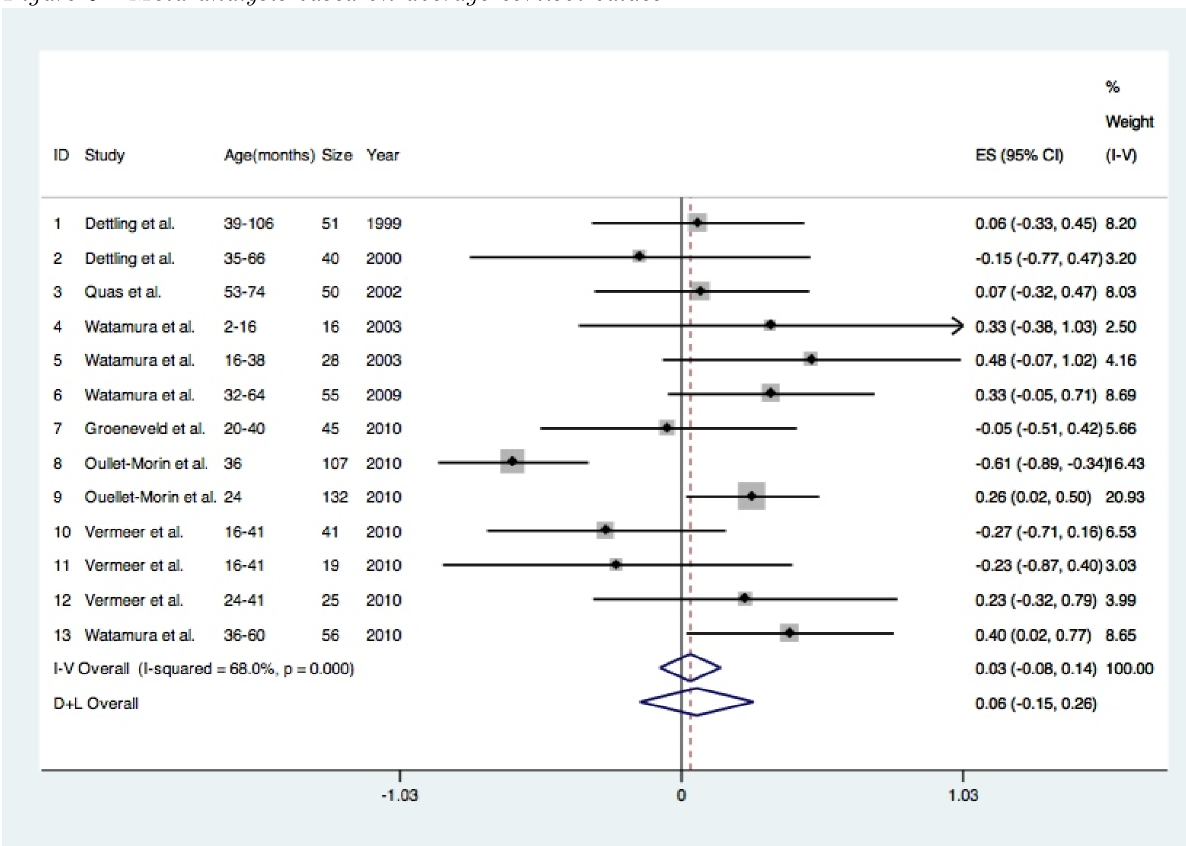


The trim and fill method yielded an effect size of d = 0.28, 95% CI [0.18, 0.38]. Compared to the unadjusted effect size calculated from the meta-analysis (d = 0.31, 95% CI [0.20, 0.41]) it represents a very slight decrease in strength, indicating no real influence of bias. A Q-test was performed to test for heterogeneity, but the result was not significant – albeit only slightly above the alpha ($\alpha = 0.05$) level ($Q = 23.52$, $df = 15$, $p = 0.074$).

3.3 Meta-Analysis Based on Average Cortisol Levels

Results from the second meta-analysis – based on compared average cortisol levels (on an absolute scale) – are presented in Figure 6.

Figure 6 - Meta-analysis based on average cortisol values



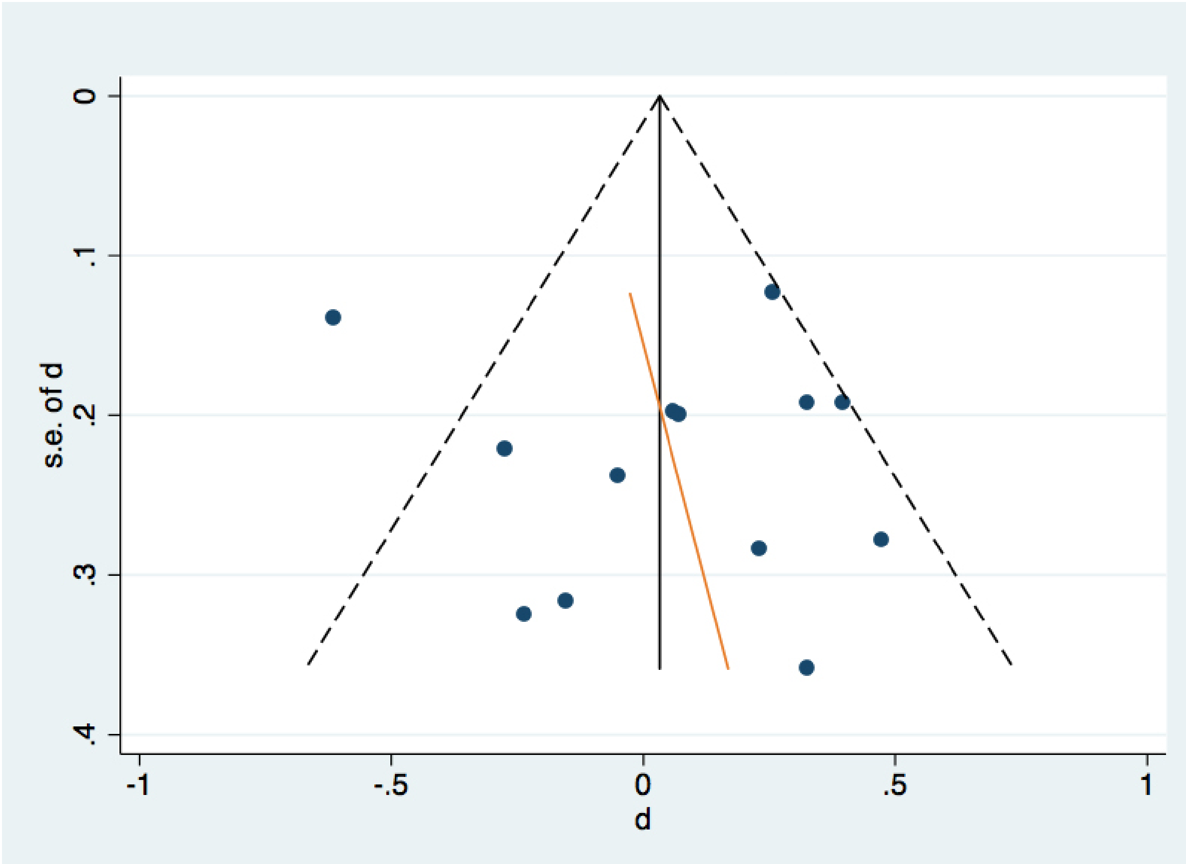
The meta-analysis was conducted on data from all available time points. (N=665)

This second meta-analysis shows that children's over-all cortisol levels were not significantly higher in childcare than at home (d = 0.06, 95% CI [-0.15, 0.26], z = 0.57, p = 0.572). This is in contrast to the first meta-analysis where childcare was shown to have a negative effect on children's diurnal cortisol patterns (d = 0.31, 95% CI [0.20, 0.41]). A chi-square was performed to test for heterogeneity and the result was significant ($\chi^2 = 37.45$, $d.f. = 12$, $p < 0.001$). Sixty-eight percent of the variation among effect sizes was attributable to heterogeneity ($I^2 = 68$).

3.3.1 Assessment of Bias

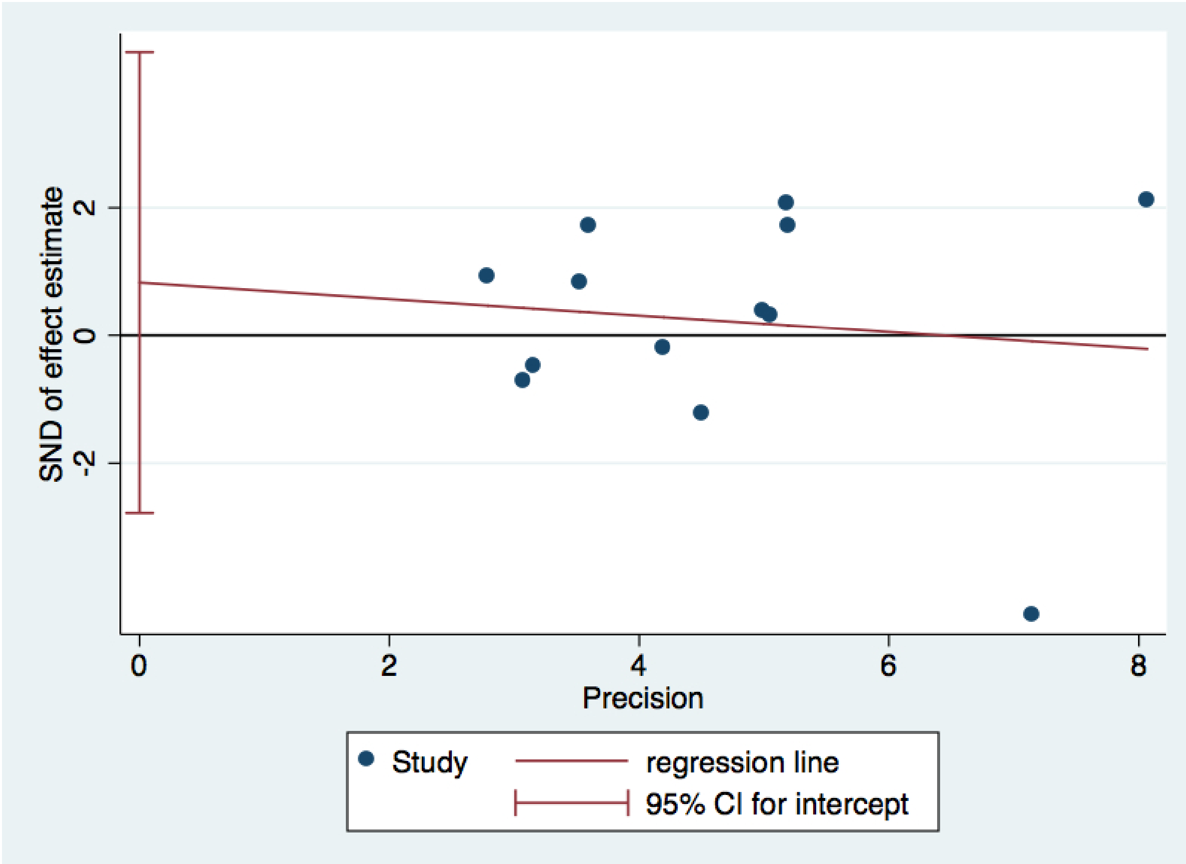
As with the first meta-analysis, bias also needs to be assessed in the meta-analysis based on average cortisol levels. The funnel plot in Figure 7 shows the sample scatter with a fitted regression line:

Figure 7 – Funnel plot with fitted regression line



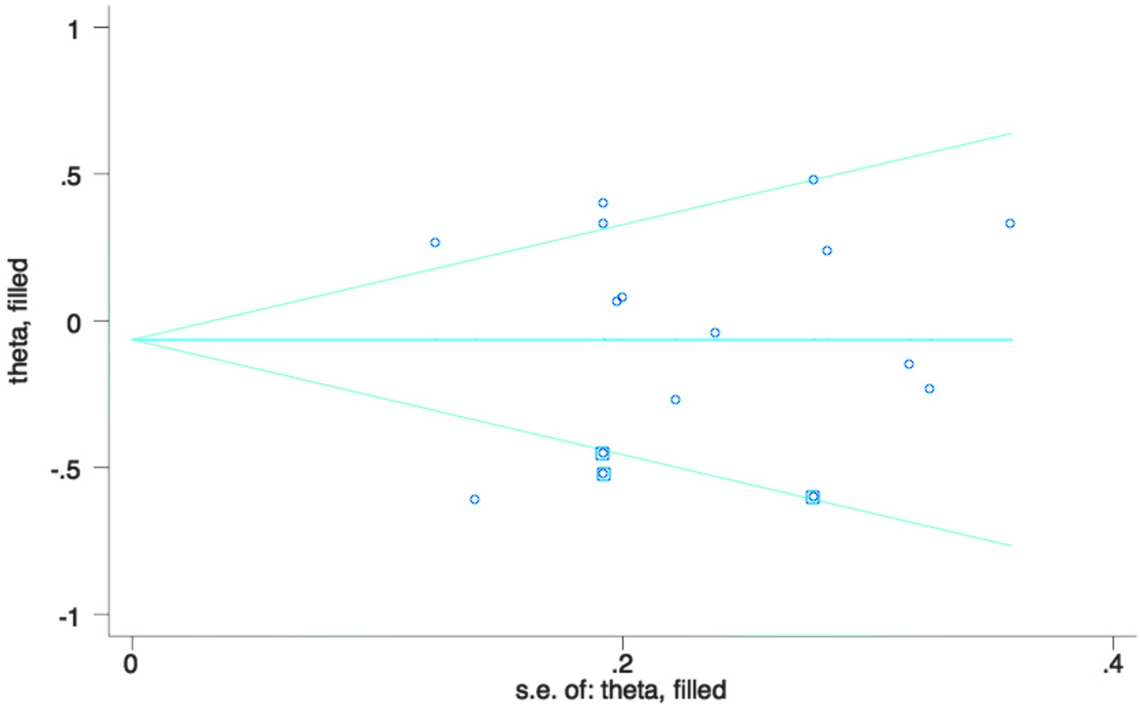
Judging by the funnel plot, the distribution of studies again seems to be slightly skewed – as is confirmed by the fitted regression line. The visual impression of the plot, in addition to the regression line does seem to suggest that the observed skew could indicate that the sample selection is biased. To quantify this potential bias, the Egger's regression test for small study effects was therefore conducted:

Figure 8 – Egger's regression test for small study effects



The Egger's regression test for small study effects estimated a slope of $\beta = -0.13$, 95% CI $[-0.87, 0.61]$, $p = 0.62$. Again, as we saw in the previous Egger's regression test, the confidence interval includes the null, rendering the test result non-significant. The null hypothesis – the hypothesis that there is no bias present – is therefore kept. Based on the Egger's regression test, the study selection does therefore not appear to be biased. The trim and fill method largely confirms this impression and is shown graphically in Figure 9.

Figure 9 – Filled funnel plot



The trim and fill method resulted in three studies being filled. The imputed studies (as indicated by the squared circles) are located on the negative side of the funnel plot, causing the summary effect estimate to be re-adjusted, but producing only a slightly lowered effect size ($d = -0.06, 95\%$ CI $[-0.17, 0.04]$) compared to the original un-trimmed meta-analysis ($d = 0.06$, 95% CI $[-0.15, 0.26]$). Lastly, a Q test was performed to test for heterogeneity, and the result was significant below the alpha ($\alpha = 0.05$) level ($Q = 54.00$, $df = 15$, $p < 0.001$).

4 Discussions

The aim of this study was to determine if childcare makes children more stressed compared to home. In total, 10 studies met the inclusion criteria and were considered to be of high quality, and were therefore reviewed and included in the meta-analyses. The results from the first meta-analysis (Figure 2) show that children in childcare display more flattened cortisol patterns than they do at home ($d = 0.31$). Although this effect was significant, the strength of the effect is arguably relatively weak (see Cohen 1988). This finding is consistent with earlier meta-analyses by Vermeer and IJzendoorn (2006) and Geoffroy, Côté, Parent & Séguin (2006), both of which estimated effect sizes based on children's diurnal cortisol patterns (with 'difference-in-difference' designs). Vermeer and IJzendoorn reported an effect size of $d = 0.37$ (equivalent to $r = 0.18$, as originally reported), while Geoffroy et al. reported an effect size of $d = 0.72$. The summary effect size from the Vermeer and IJzendoorn study lies very close to the effect size estimate computed in my meta-analysis (Figure 2), but Geoffroy et al. report a notably stronger effect than what was found in this study. The different effect sizes found in these studies may possibly be attributable to them having used different inclusion criteria, and having been conducted at different time points, with different studies available.

The results from the second meta-analysis (Figure 6) suggest that children – on an absolute scale – do not display higher cortisol levels in childcare than they do at home ($d = 0.06$). This finding is consistent with a study by Vermeer et al. (2010), who found that the cortisol levels were not significantly different for children in childcare compared to children at home.

As has been discussed earlier, results showing that the diurnal cortisol patterns were flattened or reversed do not say anything directly about the absolute levels of cortisol. Thus, a child in childcare who has on average much lower levels of cortisol in the morning than a child at home, and whose levels then do not decrease substantially, will still be interpreted as being more stressed than a child who starts off very high, but decreases more extensively by the afternoon. Therefore, including analyses on children's average levels as well, gives a more complete picture of overall stress. Thus, we can say that not only were the diurnal patterns flattened or reversed, but also that the average levels of cortisol did not differ.

However, these findings should be seen in light of this study's potential limitations. Some relate directly to methodological issues in regard to the meta-analyses. Other methodological issues are perhaps more theoretically central, and relate to limitations in the design of the studies that have been included in the meta-analyses. It is important to review limitations in both regards, since the inferences that one can make from the meta-analyses hinge upon the soundness of the original studies. I will therefore review possible limitations both with regards to the meta-analyses, and in regards to the general framework of the original studies themselves.

4.1 Limitations Related to Meta-Analysis

4.1.1 Publication bias

Publication bias – which is most commonly addressed in meta-analysis – is a term used to describe the phenomenon where the published research literature is unrepresentative or systematically different from the unpublished research literature (Rothstein, Sutton & Borenstein 2005). One of the main causes for this is that statistically significant studies, and studies with higher effect sizes are more likely to be published than studies without significant findings or low effect sizes (Rothstein et al. 2005; Borenstein et al. 2009). Although publication bias affects all areas of research, whether quantitative or qualitative, it has received considerably more attention within meta-analysis, possibly because meta-analysis has been promoted as more accurate than other approaches (Rothstein et al. 2005; Borenstein et al. 2009). It is also perhaps a matter of a cumulative effect since published studies are collected and synthesized on a larger scale within meta-analysis, and therefore – if left untreated – could undermine meta-analysis' reputation of increased accuracy and reliability.

In order to maintain the integrity of the meta-analyses, publication bias should therefore be assessed. However, in this study the results from the bias assessments did not indicate that the meta-analyses were likely to be affected by publication bias.

4.1.2 Comparing studies that are substantially different

Meta-analyses combine many different studies and generate one effect size to represent them all – like mixing apples and oranges. It is argued that in doing so, important study-level information is lost in the process (Borenstein et al. 2009). While it is the case that meta-analyses do include many different studies, Borenstein et al. assert that although it is necessary to make judgments on the appropriateness of combining studies that may vary in similarity, it is also the case that meta-analyses generally address broader questions than the individual studies themselves. In that perspective, it makes sense to combine studies across a wider range in order to achieve higher generalizability.

4.1.3 Non-comparability due to differing baseline values

As was discussed in the method section, a 'difference-in-difference' (Meyer 1995; What Works Clearinghouse 2008) design was adapted in the first meta-analysis (Figure 2) where changes in diurnal cortisol patterns were compared between the childcare group and the home group. This design allowed for a comparison between the two groups on a relative scale, in contrast to the second meta-analysis (Figure 6.) where cortisol levels were compared on an absolute scale. However, the 'difference-in-difference' design comes with a problem that was first identified by Wilder (as cited in Benjamin 1963), and formulated in his law of initial values [LIV], stating that:

> (...) the magnitude of autonomic response to a stimulus is related to the prestimulus level. (...) [N]ot only the intensity but also the direction of a response of a body function to any agent depend to a large degree on the initial level of that function at the start of the experiment. The higher this 'initial level', the smaller is the response to function-raising, the greater is the response to function-depressing agents. At more extreme initial levels there is a progressive tendency to 'no response', and to 'paradoxic reactions,' i.e. a reversal of the usual direction of response. (p. 556)

The problem becomes apparent in situations where the baseline levels (in this case, the morning levels) of the comparison groups differ. The phenomenon is best explained by an example:

Example 1

	Type of care	**Morning measure**	**Afternoon measure**
Child A	Home	4.0 nmol/L	2.0 nmol/L
Child B	Childcare	15.0 nmol/L	10.0 nmol/L

At first glance we see that child B displays markedly higher cortisol levels than child A, both at morning and in the afternoon. However, when the differences are calculated the picture becomes less clear. Child A exhibits a cortisol reduction of 2.0 nmol/L at home, while child B exhibits a reduction of 5.0 nmol/L in childcare. Both children experience a decline in cortisol throughout the day, but despite having the highest cortisol levels, child B also shows the largest decline in cortisol. On an absolute scale (as in the second meta-analysis) this example would have yielded a positive effect size, indicating that childcare is associated with elevated cortisol levels. On a relative scale (as in the first meta-analysis), however, this example would have produced a negative effect size, indicating that children in childcare are associated with stronger diurnal cortisol patterns (which is preferable, as explained in the introduction).

The problem is, as Wilder describes it, that the outcome value is dependent on the initial value. As in the example, it is therefore not unlikely that since child B displays such extreme morning levels, child B is less likely than child A to show major increases in cortisol since the levels are already elevated – even after being exposed to potential stressors in childcare. Extreme morning values may therefore instead lead to paradoxical responses since the levels are presumably less likely to stay so elevated through the whole day, despite continued exposure. This causes problems for the comparability of child A and child B since they do not have the same starting point and may therefore react very differently physiologically to the same stressors. To deal with this problem it has been suggested that by "relating values to the first principal component axis, this spurious effect can be eliminated, thus allowing of any real dependency on initial values" (For review, see Myrtek & Foerster 1985, p. 227). Despite this, statistical treatment of LIV, as proposed by Benjamin

(1963), was not implemented in the meta-analysis (Figure 2) due to cortisol data not being available on the individual level – as required for calculating the LIV corrected scores. However, the impact of LIV has been debated, where Myrtek and Foerster (1986) argue that the influence of LIV is limited, contrary to general belief, and consider it to be a rare exception rather than the rule.

4.1.4 Statistical Heterogeneity

The meta-analyses showed considerable between-study variation in effect sizes, with some studies reporting significantly higher results than other studies. As discussed in the method chapter, there are two ways of seeing this. In accordance with the random effects model, there is believed to be a variety of 'correct' results, and these might be normally distributed across a population. On the other hand, in accordance with the fixed effect model, one assumes that there is one true underlying effect and the discrepancies must be explained by other means, like random chance or measurement error. Kaitz et al. (2012) researched the nature of these between-study variations by examining if the variation in measurement results could be explained by measurement errors. Their focus was on noncompliant children, and how their resistance affected the cortisol levels. Their results showed that about 12% of the children were noncompliant at least once during sampling. Noncompliance in this case meant that a child was either reluctant to be measured, or refused entirely. Kaitz et al. (2012) concluded that children's noncompliance during saliva measurements could be source of "nonrandom missing data or extreme cortisol values" (p. 121). Kaitz et al. point to a key problem with these sampling situations. If noncomplying children are systematically different from complying children, as was suggested by Kaitz et al., then it might bias the results. Lastly, heterogeneity could also be a result of true variation in diurnal cortisol patterns. Stone et al. (2001) estimated that over 10% of a sampled population would not show the typical diurnal cortisol rhythm, indicating that there is significant between-subject variation in diurnal rhythms within a population.

4.1.5 Garbage in, garbage out

The garbage in, garbage out argument is based on the notion that synthesizing studies of poor quality will result in erroneous results (Borenstein et al. 2009). Making matters worse, a meta-analysis might further obscure this fact by falsely giving the impression of providing accurate and reliable results in situations where they are not. While the impression of meta-analysis as a purely objective endeavor might be held by some, DeCoster (2004) instead describes it as "shared subjectivity" (p. 4) since all analyses require a certain amount of subjective decisions to be made. However, these processes are always made transparent in meta-analysis (DeCoster 2004). Meta-analyses should also have clearly defined inclusion criteria that should guard against studies of poor quality from cropping up in the analyses (Borenstein et al. 2009). Borenstein et al. (2009) do however concede that even the studies that end up being included in meta-analysis will most likely not be perfect, necessitating careful bias assessment. Some of these objections relate to the internal validity of the studies that are included in the meta-analysis. Since this is crucial to the interpretation of the meta-analysis, I will review potential limitations.

4.2 Limitations Related to the Included Studies

4.2.1 Internal validity

Internal validity is described as the validity of a causal inference (Trochim 2012a). It relates to whether an observed effect is also actually caused by the predictor variable in question. In this case, internal validity relates to whether the observed changes in diurnal cortisol patterns in the childcare group is actually also caused by something within childcare itself, and not by other unknown factors. Because there are multiple factors that can threaten the internal validity of the results these will therefore be discussed in this chapter.

Selection bias When trying to determine whether childcare is making children more stressed it is necessary to have a population to study. Since studying entire populations is neither feasible nor desirable, you instead rely on studying samples of the population. These samples should ideally be representative of the rest of the population so that any findings you make will be generalizable and therefore valid for the rest of the target population. If the samples for various reasons are unrepresentative of target population then selection bias may arise (Aschengrau & Seage III 2008). Specifically, getting representative samples is achieved by selecting participants at random, a procedure known as randomization (Dietz and Kalof 2009). The opposite of selecting at random is selecting by preference, also known as convenience sampling. Convenience sampling is methodically problematic since statistical inference depends on the process of randomization in order to make valid claims to the generalizability of the results (Dietz & Kalof 2009). The majority of the studies that were included in the meta-analyses did not use randomization procedures to generate their samples, causing legitimate concern with regards to the generalizability of the results.

Attrition bias Another threat to validity is attrition. Attrition bias can occur when you have systematic loss of participants from the sample (Yakoub, Panesar & Athanasiou 2010).High refusal or dropout rates were common problems among several of the included studies. These dropouts are categorized into two categories of missing data. Donders, van der Hejden, Stijnen and Moons (2006) define the first type of missing data, random missing data, as the probability that the missing data is not linked to any individual characteristics. In such cases the randomized sample maintains its consistency, albeit with a potential loss of power. Conversely, the second type of missing data, non-random missing data, is described as the probability of the missing data being dependent on unobserved factors and therefore being inextricably linked to characteristics or circumstances specific to the individuals themselves (Donders et al 2006.). The latter case of missing data may result in a biased sample since the non-participating group may be systematically different from the participating group. Although some of the studies attempted to remedy the problem by testing for differences between participants and non-participants it is nonetheless an issue of being able to account for all relevant factors that might differ across the groups.

Ultimately, the likelihood of attrition or convenience sampling playing a major role in rendering the results non-generalizable is not easy to determine. Presumably most children are biologically fairly similar in terms of their capacity towards normal HPA functioning, something that would speak in favor of maintained generalizability. However, should there exist significant variation in diurnal cortisol naturally within populations then extra care should be given to sampling procedures and the prevention of attrition. With random sampling it would have been possible to account natural deviations, but with convenience samples it is not (Dietz & Kalof 2009). Because it is not possible to control for or to quantify the amount of error in a convenience sample, Dietz and Kalof therefore caution the application of statistical tools based on random sampling when the samples have been drawn by convenience.

Measurement bias As we have seen from the meta-analyses, there exists quite a bit of variation in effect sizes, which are reflections of the measured cortisol levels. In order to determine if such variation in cortisol measurements were real or artifacts of measurement errors, Kaitz et al. (2012) researched the phenomenon and found that noncompliance in children affected the cortisol measurements. Even though some of the included studies specified that the children were only to participate in saliva sampling voluntarily, and were under no circumstances to be coerced into providing samples, it is never the less clear that some of the sampled children may have fallen into the category defined as 'noncompliant' by Kaitz et al. Noncompliance not only referred to children who refused sampling, but also those who were somewhat reluctant or hesitant but provided saliva anyway. As such, it is therefore not unlikely that noncompliant children can be a source of measurement bias.

The use of sweeteners during saliva sampling Many of the included studies used Kool-Aid (or similar) with sweeteners to stimulate saliva flow during sample collection. However, the practice has been discouraged since studies have shown a correlation between artificial sweeteners and aberrant cortisol levels (Centre for Studies on Human Stress 2007). Additionally, Schwartz et al. (1998) found that drink mix crystals, such as Kool-Aid, artificially amplified the cortisol measurements, while Gonzalez-Bono et al. (2002) found that the acute cortisol response is largely controlled by the same centers that regulate glucose, indicating that eating prior to cortisol sampling could lead to exaggerated cortisol levels.

Confounding The aim of this study was to attempt to answer the question of whether childcare increases children's stress levels or not, but without isolating the factors that are contributing to increased cortisol levels we can only speculate that it is something within childcare itself that is causing these cortisol fluctuations. It should be emphasized that there could be other mechanisms at play, accounting for some (or all) of the differences between the childcare measurements and the home measurements. When not all factors have been accounted for, a spurious association between the exposure and the outcome can occur, which is commonly known as confounding (Bonita, Beaglehole & Kjellström 2006). Confounding can occur when another exposure variable is present which is correlated with both the exposure and the outcome (Bonita, Beaglehole & Kjellström 2006). Confounders can be hard to control, but are vital to be aware of in order to find out what is really causing cortisol levels to elevate. I will therefore address some possible confounders in the following section.

Day of sampling Considering that the majority of the studies opted for within-subject designs where the same children were measured both in childcare and at home, the question of when to collect home samples naturally arises. Since the children are mainly in childcare during the weekdays, home samples would have to be collected during weekends, which was the case for most of the included studies. However, it is not unlikely that for many children the weekends are associated with quality time with the families, trips and fun activities, and perhaps also rest. Conversely, children living under poor home conditions might associate the weekends with something negative, an absence of care and the safe boundaries provided by childcare. In either case, one cannot rule out that these factors could have had an unintended effect on the cortisol measurements.

Socioeconomic status Socioeconomic status is a possible confounder because it potentially affects both the likelihood of being in childcare and a child's resilience or robustness in relation to stress coping (Bradly & Corwyn 2002). Norway has a relatively high degree of social equality (Norwegian Ministry of Labour and Social Inclusion 2011) which is likely a contributing factor for why we see that it is the children from high socioeconomic backgrounds that are the most likely to attend childcare in Scandinavian countries (Hjern, Haglund, Rasmussen & Rosén 2000). However, for countries with low socioeconomic equality, like USA (which is represented by 33% of the sampled children in the meta-analyses) it is presumably the other way around, with children in families with low socioeconomic status being the most likely to attend childcare. Socioeconomic status [SES] can therefore be a possible confounder since research has shown that SES is associated with health outcomes, and that children from low socioeconomic backgrounds are more exposed to stress-inducing home environments, which in turn could result in children with higher reactivity to stressors (Bradley & Corwyn 2002).

Illness Illness has been shown to increase cortisol production in the body (Centre for Studies on Human Stress 2007), which is why several studies attempted to control for illness by instructing parents to not collect saliva samples on days when the children were sick. Illness could nevertheless be a confounder that is hard to control for, as it is probably tied to socioeconomic status, in that families with low SES are presumably more likely to send their children to childcare when sick than families with high SES. As a consequence of this, the overall cortisol levels in the childcare group increase, biasing the results.

Heterogeneity It is possible that childcare may cause stress in some groups of children, but not all. However, if these groups are analyzed as a whole, one might not discover the true effect within the subgroup. A subgroup analysis is defined as a "statistical analysis that establishes [an] effect (...) with one or more subgroups of the sample defined by baseline characteristics" (Malani, Bembom & van der Laan 2008, p. 2). Such analyses can be vital to determining whether there are substantial differences between a specific part of a population and the population as a whole. For instance, stratifying the sample by 'children in high quality daycare' and 'children in low quality daycare,' could provide important information with regards to understanding whether quality of childcare is correlated with children's stress levels. We know from earlier research (Melhuish 2004), that high quality daycare can affect certain subgroups of children differently, depending on whether they come from so-called disadvantaged families or not. The same might be true for stress levels, since certain children might be more susceptible to stress than others, given certain conditions.

In the current study, subgroup analyses were nevertheless not performed due to the relatively limited number of studies included in the meta-analyses, and the lack of consistent classification criteria (e.g., quality of childcare) across the different studies.

4.2.2 External validity

External validity relates to the generalizability of the results to other populations (Trochim 2012b). If similar results have been found in studies of other populations then the external validity of the findings are strengthened (Bonita, Beaglehole & Kjellström 2006). In this case, the main aim was to find out if children in childcare have higher cortisol levels than children at home, and to quantify the effect. In that regard, a representative selection would be the ideal. That would be achieved by including a wide selection of children from different socioeconomic backgrounds, attending both low quality and high quality childcare, living in rural and urban areas, etc. While such diversity serves as a good representation of what the average child may experience, the average effect could also conceal real differences in effects between subgroups. It could naturally be the case that analyzing a subgroup sample might yield the same results as the entire sample, and that children do not systematically differ in stress-responses depending on age, gender, background, etc., but without doing subgroup analyses there is no way of being certain.

5 Conclusion

In summary, this study replicated – to some degree – earlier findings suggesting that children's diurnal cortisol patterns tend to be more flattened in childcare than at home. However, the effect was weaker than what have been found in earlier meta-analyses, and it is unclear if the effect is of a magnitude that is likely to have an impact on health outcomes. Secondly, I found no evidence to support the notion that children in childcare have higher cortisol levels on average (on an absolute scale), giving further support of minimal effects of stress related to childcare. However, as noted in the discussion, it is important to keep in mind that these are aggregate effects, and that there could be significant effects found on an individual or subgroup level. It should also be clear from the discussions that there are numerous factors contributing to an element of uncertainty related to the results, which emphasizes the need for more rigorous study designs in future research. Although the meta-analyses performed in this study revealed no major effects of stress in childcare, it is nonetheless clear from the studies examined that there are worrisome aspects of childcare that should be further investigated. Larger, and national studies should be conducted to determine with greater certainty what groups are at risk and how the effects can be moderated. Considering that populations differ in terms of socioeconomic distribution, childcare models, quality of care, etc., it is therefore unclear how the present findings translate to the general childcare population of Norway. The conducting of national studies is therefore encouraged.

While the biological implications of high stress levels are certainly important to investigate, as have been the focus of this study, due attention should also be given to the subjective aspects of stress. Further research should therefore attempt to investigate how the children themselves perceive the effects of elevated cortisol levels, and to which degree such elevations cause experiences of felt stress or whether it triggers stressful behavior.

If we agree with Borenstein et al. (2009), that the purpose of the meta-analysis is ultimately to inform policy then it raises the question of what such a reform should entail. Despite some public concern (see Tveitereid 2008), childcare is undoubtedly here to stay, and is serving an important function as a pedagogical institution, while simultaneously contributing to higher social equality (Havnes & Mogstad 2011). Perhaps, then, the question is not so much whether children should be in childcare, but rather how childcare services may be tailored in a larger degree to children's specific needs in order to reduce stress. Ultimately, childcare ecologies are complex and should be viewed holistically when evaluating the pros and cons, as other positive aspects might outweigh stress in terms of overall benefit.

References

Adam, E. K., & Kumari, M. (2009). Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*, *34*, 1423-1436.

Aschengrau, A., & Seage, G. R. (2008). *Essentials of epidemiology in public health.* USA: Jones and Bartlett Publishers, inc.

Benjamin, L. S. (1963). Statistical treatment of the law of initial values (LIV) in autonomic research: A review and recommendation. *Psychosomatic Medicine*, *25*, 556-566.

Bonita, R., Beaglehole, R., & Kjellstrom, T. (2006). *Basic epidemiology, 2nd edition.* Switzerland: World Health Organization.

Borenstein, M. (2005). Software for publication bias. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis - prevention, assessment and adjustments* (p. 194-220). England: John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* United Kingdom: John Wiley & Sons Ltd.

Borg, E., Kristiansen, I.-H., & Backe-Hansen, E. (2008). *NOVA rapport 06/08: Kvalitet og innhold i norske barnehager – en kunnskapsoversikt.* Oslo: Norwegian Social Research (NOVA).

Bowlby, J. (1977). The making and breaking of affectional bonds. I. Aetiology and psychopathology in light of attachment theory. *The British Journal of Psychiatry*, *130*, 201-210.

Bradly, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371-399.

Brunner, E. J., Hemingway, H., Walker, B. R., Page, M., Clarke, P., Juneja, M., ... Marmot, M. G. (2002). Adrenocortical, autonomic, and inflammatory causes of the metabolic syndrom. *Circulation*, *106*, 2659-2665.

Centre for Studies on Human Stress. (2007). *How to measure stress in humans?* Retrieved from http://www.humanstress.ca/documents/pdf/Mesures%20physiologiques/CESH_howMesureStress-MB.pdf

Chrousos, G. P. (1998). Stressors, stress, and neuroendocrine integration of the adaptive response - the 1997 Hans Selye memorial lecture. *Annals of the New York Academy of Sciences*, *851*, 311–335.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* USA: Lawrence Erlbaum Associates.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, *24*, 385-396.

Cohen, S., Kessler, R. C., & Gordon, L. U. (1995). *Measuring stress – a guide for health and social scientists.* UK: Oxford University Press.

DeCoster, J. (2004). *Meta-analysis notes.* Retrieved from http://www.stat-help.com/meta.pdf

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2006). Analysing and presenting results. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions - version 4.2.6* (p. 97-166). The Cochrane Collaboration. Retrieved from http://www.cochrane.org/sites/default/files/uploads/Handbook4.2.6Sep2006.pdf

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2011). Analysing data and undertaking meta-analysis. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions - version 5.1.0* (p. 97-166). The Cochrane Collaboration.

de Haan, M., Gunnar, M. R., Tout, K., Hart, J., & Stansbury, K. (1998). Familiar and novel contexts yield different associations between cortisol and behavior among 2-year-old children. *Dev Psychobiol*, *33*, 93-101.

Dettling, A. C., Gunnar, M. R., & Donzella, B. (1999). Cortisol levels of young children in full-day childcare centers - relations with age and temperament. *Psychoneuroendocrinology*, *24*, 519-536.

Dettling, A. C., Parker, S. W., Lane, S., Sebanc, A., & Gunnar, M. R. (2000). Quality of care and temperament determine changes in cortisol concentrations over the day for young children in childcare. *Psychoneuroendocrinology*, *25*, 819-836.

de Weerth, C., Zijl, R. H., & Buitelaar, J. (2003). Development of cortisol circadian rhythm in infancy. *Early Human Development*, *73*, 39-52.

Dietz, T., & Kalof, L. (2009). *Introduction to social statistics - the logic of statistical reasoning.* UK: Wiley-Blackwell.

Donders, A. R. T., van der Hejden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review – a gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology*, *59*, 1087-1091.

Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis – prevention, assessment and adjustments.* UK: John Wiley & Sons.

Evans, G. W. (2006). Child development and the physical environment. *Annu. Rev. Psychol.*, *57*, 423-451.

Geoffroy, M. C., Cote, S. M., Parent, S., & Seguin, J. R. (2006). Daycare attendance, stress, and mental health. *Canadian journal of psychiatry*, *51*(9), 607-615.

Glasziou, P., Irwig, L., Bain, C., & Colditz, G. (2001). *Systematic reviews in health care: A practical guide.* USA: Cambridge University Press.

Gonzalez-Bono, E., Rohleder, N., Hellhammer, D. H., Salvador, A., & Kirschbaum, C. (2002). Glucose but not protein or fat load amplifies the cortisol response to psychosocial stress. *Hormones and Behavior*, *41*, 328-333.

Groeneveld, M. G., Vermeer, H. J., van IJzendoorn, M. H., & Linting, M. (2010). Children's wellbeing and cortisol levels in home-based and center-based childcare. *Early Childhood Research Quarterly*, *25*, 502-514.

Gulbrandsen, L. (2009). *Førskolelærere og barnehageansatte.* Oslo: Norwegian Social Research (NOVA).

Gulbrandsen, L., Johansson, J.-E., & Nilsen, R. D. (2002). *Forskning om barnehager - en kunnskapsstatus.* Oslo: The Research Council of Norway.

Gunnar, M. R., Kryzer, E., van Ryzin, M. J., & Phillips, D. A. (2010). The rise in cortisol in family day care - associations with aspects of care quality, child behavior, and child sex. *Child Development*, *81*(3), 851-869.

Gunnar, M. R., Kryzer, E., van Ryzin, M. J., & Phillips, D. A. (2011). The import of the cortisol rise in child care differs as a function of behavioral inhibition. *Developmental Psychology*, *47*(3), 792-803.

Halligan, S. L., Herbert, J., Goodyer, I. M., & Murray, L. (2004). Exposure to postnatal depression predicts cortisol in adolescent offspring. *Biol Psychiatry*, *55*, 376-381.

Hardeland, R. (2000). An ancient cellular mechanism in the tension field between energy requirements and destruction avoidance. In T. V. Driessche, J.-L. Guisset, & G. M. Petiau-de Vries (Eds.), *The redox state and circadian rhythms.* The Netherlands: Kluwer Academic Publishers.

Havnes, T., & Mogstad, M. (2011). No child left behind – subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy*, *3*(2), 97-129.

Herbert, J., Goodyer, I. M., Grossman, A. B., Hastings, M. H., de Kloet, E. R., Lightman, S. L., ... Seckl, J. R. (2006). Do corticosteroids damage the brain? *Journal of Neuroendocrinology*, *18*, 393-411.

Hjern, A., Haglund, B., Rasmussen, F., & Rosen, M. (2000). Socio-economic differences in daycare arrangements and use of medical care and antibiotics in swedish preschool children. *Acta Paediatrica*, *89*, 1250-1256.

Kaitz, M., Sabato, R., Shalev, I., Ebstein, R., & Mankuta, D. (2012). Children's noncompliance during saliva collection predicts measures of salivary cortisol. *Dev Psychobiol*, *54*, 113-123.

Kirschbaum, C., & Hellhammer, D. H. (1994). Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. *Psychoneuroendocrinology*, *19*(4), 313-333.

Klein, P. S., Kraft, R. R., & Shohet, C. (2010). Behaviour patterns in daily mother-child separations – possible opportunities for stress reduction. *Early Child Development and Care*, *180*(3), 387-396.

Legendre, A. (2003). Environmental features influencing toddlers' bioemotional reactions in day care centers. *Environment and Behavior*, *35*(4), 523-549.

Li, I., & Shen, P.-S. (2008). Internalizing disposition and preschool children's cortisol fluctuations. *Child: care, health and development*, *34*(5), 626-630.

Lundberg, U., Westermark, O., & Rasch, B. (1993). Cardiovascular and neuroendocrine activity in preschool children - comparison between day-care and home levels. *Scandinavian Journal of Pscyhology*, *34*(371-378).

Malani, A., Bembom, O., & van der Laan, M. (2008). *Reforming subgroup analysis.* Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1119970

Melgård, M., & Larsen, S. G. (2008, 28. August). *-Barnehage kan være skadelig for barna.* Dagbladet. Retrieved from http://www.dagbladet.no/nyheter/2008/08/29/545166.html

Melhuish, E. C. (2004). *A literature review of the impact of early years provision on young children, with emphasis given to children from disadvantaged backgrounds - report to the comptroller and auditor*

general. London: National Audit Office.
Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*, *13*(2), 151-161.
Myrtek, M., & Foerster, F. (1985). The law of initial value - a rare exception. *Biological Psychology*, *22*, 227-237.
Norwegian Labor Inspection Authority. (2005). *Sluttrapport, europeisk arbeidsmiljøuke 2005 - støy på arbeidsplassen.* Retrieved from http://www.arbeidstilsynet.no/binfil/download2.php?tid=91537
Norwegian Ministry of Education and Research. (2005). *Act no. 64 of june 2005 relating to kindergartens (the kindergarten act).* Retrieved from http://www.regjeringen.no/upload/KD/Vedlegg/Barnehager/engelsk/Act_no_64_of_June_2005_web.pdf
Norwegian Ministry of Education and Research. (2006). *Framework plan for the content and tasks of kindergartens.* Retrieved from http://www.regjeringen.no/upload/kilde/kd/reg/2006/0036/ddd/pdfv/285775-rammeplanen-engelsk-pdf.pdf
Norwegian Ministry of Education and Research. (2008). *Rett til barnehageplass for alle barn.* Retrieved from http://www.regjeringen.no/en/dep/kd/press-contacts/Press-releases/2008/rett-til-barnehageplass-for-alle-barn.html?id=509127
Norwegian Ministry of Education and Research. (2008-2009). *Stortingsmelding nr. 41 – kvalitet i barnehagen.* Retrieved from http://www.regjeringen.no/pages/2197014/PDFS/STM200820090041000DDDPDFS.pdf
Norwegian Ministry of Labour and Social Inclusion. (2011). *Nou 2011:5 – grunnlaget for inntektsoppgjorene 2011.* Retrieved from http://www.regjeringen.no/pages/16468487/PDFS/NOU201120110005000DDDPDFS.pdf
Office of the Prime Minister. (2005). *Platform for regjeringssamarbeidet mellom arbeiderpartiet, sosialistisk venstreparti og senterpartiet.* Retrieved from http://www.regjeringen.no/upload/SMK/Vedlegg/2005/regjeringsplatform_SoriaMoria.pdf
Olsø, E. F. (2011, 04. July). *Ettåringer utslitte av barnehage.* Adressa. Retrieved from http://www.adressa.no/nyheter/trondheim/article1656724.ece
Ouellet-Morin, I., Tremblay, R. E., Boivin, R. E., Meaney, M., Kramer, M., & Cote, S. M. (2010). Diurnal cortisol secretion at home and in child care - a prospective study of 2-year-old toddlers. *Journal of Child Psychology and Psychiatry*, *51*(3), 295-303.
Quas, J. A., Murowchick, E., Bensadoun, J., & Boyce, T. W. (2002). Predictors of children's cortisol activation during the transition to kindergarten. *Developmental and behavioral pediatrics*, *23*(5), 304-313.
Rappolt-Schlichtmann, G., Willet, J. B., Ayoub, C. C., Lindsley, R., Hulette, A. C., & Fischer, K. (2009). Poverty, relationship conflict, and the regulation of cortisol in small and large group contexts at child care. *The International Mind, Brain, and Education Society*, *3*(3), 131-142.
Rodriguez, C., Mayo, J. C., & Sainz, R. M. (2000). Apoptotic signals - possible implications of circadian rhythms. In T. V. Driessche, J.-L. Guisset, & G. M. Petiau-de Vries (Eds.), *The redox state and circadian rhythms* (p. 1-4). The Netherlands: Kluwer Academic Publishers.
Rothstein, H., Sutton, A. J., & Borenstein, M. (2005). Publication bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis – prevention, assessment and adjustments* (p. 1-7). England: John Wiley & Sons.
Sæther, J. P. (2010). *Barn i barnehage - foreldrebakgrunn og utvikling de seneste årene.* Oslo: Statistics Norway.
Schwartz, E. B., Granger, D. A., Susman, E. J., Gunnar, M. R., & Laird, B. (1998). Assessing salivary cortisol in studies of child development. *Child Development*, *69*(6), 1503-1513.
Sims, M., Guilfoyle, A., & Parry, T. S. (2006). Children's cortisol levels and quality of child care provision. *Child: Care, Health and Development*, *32*(4), 453-466.
Sørensen, F. (2011, 06. July). *Avviser stressmåling av de minste.* Retrieved from http://www.barnehage.no/no/Nyheter/2011/Juli/Avviser-stressmaling-av-de-minste/
Statistics Norway. (2012a). *Barn i barnehager etter alder og fylke - 2011.* Retrieved from http://www.ssb.no/emner/04/02/10/barnehager/tab-2012-06-15-02.html
Statistics Norway. (2012b). *Arbeidsliv, yrkesdeltaking og lønn.* Retrieved from http://statbank.ssb.no/statistikkbanken/Default_FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selectvarval/define.asp&Tabellid=08720

Statistics Norway. (2012c). *Table 04683 - c. barnehager - nivå 3 (k).* Retrieved from http://statbank.ssb.no/statistikkbanken/selectvarval/Define.asp?subjectcode=&ProductId=&MainTable=Kostra3KCBarneha&nvl=true&PLanguage=0&nyTmpVar=true

Statistics Norway. (2012d). *Table 07459 - folkemengde, etter kjønn og ettårig alder. 1. januar (k).* Retrieved from http://statbank.ssb.no/statistikkbanken/selectvarval/Define.asp?subjectcode=&ProductId=&MainTable=NY3026&nvl=true&PLanguage=0&nyTmpVar=true

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis – prevention, assessment and adjustments* (p. 99-110). England: John Wiley & Sons.

Sterne, J. A. C., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal*, *4*(2), 127-141.

Stone, A. A., Schwartz, J. E., Smyth, J., Kirschbaum, C., Cohen, S., Hellhammer, D. H., & Grossman, S. (2001). Individual differences in the diurnal cycle of salivary free cortisol: A replication of flattened cycles in some individual. *Psychoneuroendocrinology*, *26*, 295-306.

Sumner, M. M., Bernard, K., & Dozier, M. (2010). Young children's full-day patterns of cortisol production on child care days. *Arch Pediatr Adolesc Med*, *164*(6), 567-571.

Tout, K., de Haan, M., Campbell, E. K., & Gunnar, M. R. (1998). Social behavior correlates of cortisol activity in child care - gender differences and time-of-day effects. *Child Development*, *69*(5), 1247-1262.

Trochim, W. M. (2012a, 24. July). *The research methods knowledge base, 2nd edition.* Retrieved from http://www.socialresearchmethods.net/kb/intval.php

Trochim, W. M. (2012b, 25. July). *The research methods knowledge base, 2nd edition.* Retrieved from http://www.socialresearchmethods.net/kb/external.php

Tveitereid, S. (2008, 29. August). *Angst i barnehagen.* Retrieved from http://www.dagbladet.no/kultur/2008/08/29/545171.html

Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* USA: University of Wisconsin-Madison, Institute for Research on Poverty.

Vassenden, A., Thygesen, J., Brosvik, S. B., Alvestad, M., & Abrahamsen, G. (2011). *Barnehagens organisering og strukturelle faktorers betydning for kvalitet – rapport iris 2011/029.* Stavanger: International Research Institute of Stavanger [IRIS].

Vermeer, H. J., Groeneveld, M. G., Larrea, I., van IJzendoorn, M. H., Barandiaran, A., & Linting, M. (2010). Child care quality and children's cortisol in basque country and the netherlands. *Journal of Applied Developmental Psychology*, *41*, 339-347.

Vermeer, H. J., & van IJzendoorn, M. H. (2006). Children's elevated cortisol levels at daycare - a review and meta-analysis. *Early Childhood Research Quarterly*, *21*, 390-401.

Watamura, S. E., Coe, C. L., Laudenslager, M. L., & Robertson, S. S. (2010). Child care setting affects salivary cortisol and antibody secretion in young children. *Psychoneuroendocrinology*, *35*, 1156-1166.

Watamura, S. E., Donzella, B., Alwin, J., & Gunnar, M. R. (2003). Morning-to-afternoon increases in cortisol concentrations for infants and toddlers at child care - age differences and behavioral correlates. *Child Development*, *74*(4), 1006-1020.

Watamura, S. E., Donzella, B., Kertes, D. A., & Gunnar, M. R. (2004). Developmental changes in baseline cortisol activity in early childhood: Relations with napping and effortful control. *Developmental Psychobiology*, *45*, 125-133.

Watamura, S. E., Kryzer, E. M., & Robertson, S. S. (2009). Cortisol patterns at home and child care - afternoon differences and evening recovery in children attending very high quality full-day center-based day care. *Journal of Applied Developmental Psychology*, *30*, 475-485.

Watamura, S. E., Sebanc, A. M., & Gunnar, M. R. (2001). Rising cortisol at childcare - relations with nap, rest, and temperament. *Developmental Psychobiology*, *40*, 33-42.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited - a meta-analysis of treatment outcome studies. *Psychological Bulletin*, *117*(3), 450-468.

What Works Clearinghouse. (2008). *Procedures and standards handbook (version 2.0).* Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v2_standards_handbook.pdf

Wilson, A. M., Sims, E. J., McFarlane, L. C., & Lipworth, B. J. (1998). Effects of intranasal corticosteroids on adrenal, bone, and blood markers of systemic activity in allergic rhinitis. *Journal of Allergy and Clinical Immunology*, *102*(4), 598-604.

Winsvold, A., & Gulbrandsen, L. (2009). *Kvalitet og kvantitet – kvalitet i en barnehagesektor i sterk vekst.* Oslo: Norwegian Social Research (NOVA).

Yakoub, D., Panesar, S. S., & Athanasiou, T. (2010). Methodological framework for evaluation and prevention of publication bias in surgical studies. In T. Athanasiou, H. Debas, & A. Darzi (Eds.), *Key topics in surgical research and methodology.* New York: Springer-Verlag Berlin Heidelberg.

Appendices

A Data used in Meta-Analysis Based on Change in Diurnal Cortisol Patterns

Table A - Data generated from calculations based on change in diurnal cortisol patterns

	Study	**Size**	**Year**	**Country**	***d***	**var_*d***	**se_*d***
1.	Dettling et al.	51	1999	Switzerland	.5158607	.0405202	.2012962
2.	Dettling et al.	40	2000	USA	.3959002	.1022098	.3197027
3.	Quas et al.	50	2002	USA	.0738717	.0400273	.2000682
4.	Legendre	73	2003	France / Hungary	.5916127	.0285959	.1691032
5.	Watamura et al.	16	2003	USA	-.0360179	.1269427	.3562903
6.	Watamura et al.	28	2003	USA	.6384484	.0789724	.2810203
7.	Watamura et al.	55	2009	USA	.5254707	.0372549	.1930154
8.	Groeneveld et al.	45	2010	The Netherlands	.1243811	.0569041	.2385459
9.	Oullet-Morin et al.	107	2010	Canada	.2133731	.0185416	.1361676
10.	Oullet-Morin et al.	132	2010	Canada	.1717711	.0148729	.1219545
11.	Vermeer et al.	41	2010	Basque Country	-.1677772	.0489521	.2212513
12.	Vermeer et al.	19	2010	Basque Country	.7931252	.1135401	.3369571
13.	Vermeer et al.	25	2010	The Netherlands	.2149014	.0804618	.2836579
14.	Watamura et al.	56	2010	USA	.5290541	.0377079	.1941852

Study = Main author of the research study.
Size = Sampled group size.
Year = Publication year.
Country = The geographical location of the daycare center.
d = Effect size, measured in Cohen's d.
var_d = Variance of d.
se_d = Standard error of d.

B Data used in Meta-Analysis Based on Average Cortisol Levels

Table B - Data generated from calculations based on average cortisol values (on an absolute scale)

	Study	**Size**	**Year**	**Country**	***d***	**var_*d***	**se_*d***
1	Dettling et al.	51	1999	Switzerland	.0604674	.0392336	.1980748
2	Dettling et al.	40	2000	USA	-.1509548	.1005355	.3170733
3	Quas et al.	50	2002	USA	.0738717	.0400273	.2000682
4	Watamura et al.	16	2003	USA	.3263513	.1285368	.3585203
5.	Watamura et al.	28	2003	USA	.4757453	.0773822	.2781766
6.	Watamura et al.	55	2009	USA	.328205	.0370224	.1924121
7.	Groeneveld et al.	45	2010	The Netherlands	-.04729	.0568306	.2383917
8.	Oullet-Morin et al.	107	2010	Canada	-.6139262	.0195738	.1399065
9.	Ouellet-Morin et al.	132	2010	Canada	.2601386	.0153671	.1239643
10.	Vermeer et al.	41	2010	Basque Country	-.2741219	.0492387	.2218979
11.	Vermeer et al.	19	2010	Basque Country	-.2339109	.1059831	.3255504
12.	Vermeer et al.	25	2010	The Netherlands	.2329884	.0805428	.2838007
13.	Watamura et al.	56	2010	USA	.3967906	.0371612	.1927724

Study = Main author of the research study.
Size = Sampled group size.
Year = Publication year.
Country = The geographical location of the daycare center.
d = Effect size, measured in Cohen's d.
var_d = Variance of d.
se_d = Standard error of d.