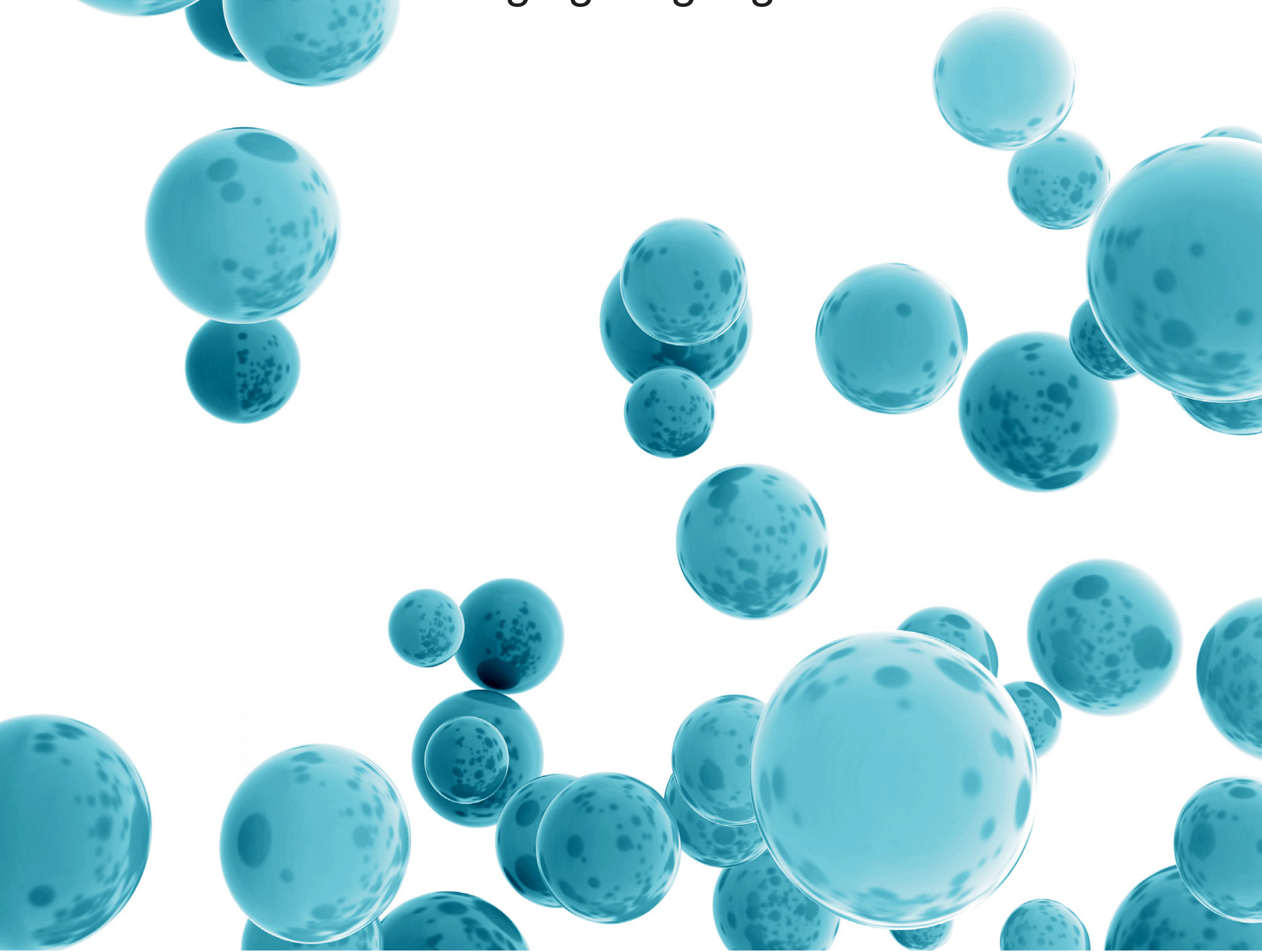
Acta Universitatis Tamperensis 2240

HANNA ELO

# Acquiring Language as a Twin

## Twin children´s early health, social environment and emerging language skills



HANNA ELO

# Acquiring Language as a Twin

## Twin children´s early health, social environment and emerging language skills

■

ACADEMIC DISSERTATION
To be presented, with the permission of
the Board of the School of Social Sciences and Humanities
of the University of Tampere,
for public discussion in the Väinö Linna auditorium K104,
Kalevantie 5, Tampere,
on 13 December 2016, at 12 o'clock.

UNIVERSITY OF TAMPERE

HANNA ELO

# Acquiring Language as a Twin

## Twin children´s early health, social environment and emerging language skills

*Acta Universitatis Tamperensis 2240*
*Tampere University Press*
*Tampere 2016*

ACADEMIC DISSERTATION
University of Tampere
School of Social Sciences and Humanities
Finland

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.



Cover design by
Mikko Reinikka

Acta Universitatis Tamperensis 2240
ISBN 978-952-03-0295-5 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1740
ISBN 978-952-03-0296-2 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

Suomen Yliopistopaino Oy – Juvenes Print
Tampere 2016

To my family

# Abstract


Typically, children learn to master the language they are exposed to in everyday interactions, but twin children are at risk for delayed language development. The onset of the delay is, however, not known, and there is no consensus about the etiology of the delay. Some studies have emphasized the role of pre- and perinatal health factors, while others underline the role of social environment and specifically the everyday interaction occurring in family homes.

Previous studies on twin family interaction have relied on small, qualitative, and non-representative samples. Therefore, in this study, a novel automated method (LENA™) was assessed, and its analyses were utilized to quantify the interaction occurring in family homes. In addition to method testing, the current study was aimed at describing the early language development of twins and studying the effects of the biomedical and social environment on twins' early language development. The developmental information of twins' language acquisition was gathered via parent reports on the onset of vocal milestones and the emergence of first words. In addition, parents reported on the development of their children's vocabularies and other language skills at the age of 12, 18, and 24 months.

The automated method showed to be reliable in detecting the speech of children and female adults, but reliable to a lesser extent in detecting male adult speech. In addition, the automated calculations turned out to be reliable for the amount of child vocalizations, but not for adult words. As for the child's development, the main results are as follows: 1) the onset of variegated babbling was substantially delayed in twins, and their lexicon size and language scores remained lower than those of children in the normative data, although remained within normal variation. 2) Older siblings influenced family interaction with their own production, but also by activating fathers and reducing the time their infant siblings vocalize. 3) Twins with older siblings showed better language skills at the age of two years than first-born twins.

Based on the results, this study suggests that the reliability and validity of the LENA System needs to be further evaluated before it can be applied to clinical use in Finnish. In addition, the delayed onset of variegated babbling and its possible relations to, for example, later phonological development should be studied as well as the enhancing role of an older sibling.

# Tiivistelmä


Tyypillisesti kehittyvät lapset omaksuvat kielen jokapäiväisissä vuorovaikutustilanteissa, mutta kaksoslasten kielen kehitys on yksöslapsia useammin viivästynyttä. Viiveen alkamisajankohtaa ei tiedetä eikä viiveen aiheuttajista ole yhdenmukaista näkemystä. Viiveen syiksi on esitetty raskauden aikaisia, synnytykseen ja vastasyntyneisyyskauteen liittyviä terveydellisiä tekijöitä, mutta myös jokapäiväisten vuorovaikutuskokemusten erilaisuutta. Tämän vuoksi tässä tutkimuksessa tarkastellaan sekä kaksoslasten varhaista kehitystä että siihen mahdollisesti vaikuttavia terveydellisiä ja sosiaalisia tekijöitä.

Aikaisemmat kaksoslasten perheympäristön ja perheen sisäisen vuorovaikutuksen tutkimukset on toteutettu pienillä näytteillä ja laadullisin menetelmin. Tässä tutkimuksessa vuorovaikutuksen tutkimusmenetelmäksi on valittu ääniaineiston automaattinen analyysimenetelmä, LENA™. LENA-menetelmä hyödyntää puhujantunnistusta ja englanninkielisellä kieliaineistolla opetettuja algoritmeja. Koska LENA-menetelmää ei ole aiemmin käytetty suomenkielisellä aineistolla, tutkimuksessa sekä arvioidaan sen luotettavuutta että hyödynnetään siitä saatavia analyyseja. Lasten kehityksellinen tieto kerättiin vanhemmilta standardoiduilla ja normeeratuilla lomakkeilla, joita käyttävät kliinistä työtä tekevät puheterapeutit ja psykologit.

Tulokset osoittavat, että LENA tunnistaa lapsi- ja naispuhujat aineistosta hyvin, mutta se ei ole yhtä tarkka miespuhujien tunnistamisessa. LENA laskee luotettavasti lasten puheenkaltaiset ääntelyt, mutta aikuisten sanamäärien laskennassa ohjelma ei vaikuta luotettavalta. Lasten kehityksen seurannan päätulokset ovat seuraavat: 1) Kaksosten varioiva jokeltelu alkoi huomattavan viiveisesti, mutta sanasto ja kielelliset taidot kehittyivät normaalivariaation sisällä. 2) Kaksoset kuulivat enemmän sisaruspuhetta perheissä, joissa oli kaksosten lisäksi vanhempia sisaruksia. Näissä perheissä isät puhuivat enemmän, mutta kaksoset ääntelivät vähemmän. 3) Kaksosilla, joilla oli vanhempia sisaruksia, oli kahden vuoden iässä suurempi sanasto ja paremmat morfosyntaktiset taidot kuin esikoiskaksosilla.

Tulokset osoittavat, että LENA-menetelmää tulisi arvioida tarkemmin ennen kuin se voidaan ottaa käyttöön kliiniseen työhön Suomessa. Lisäksi tulosten perusteella näyttäisi siltä, että sekä varioivan jokeltelun viivästymistä kaksosilla että vanhemman sisaruksen rikastavaa merkitystä nuorempien sisarusten kielen kehitykselle tulisi tarkastella lisää myöhemmissä tutkimuksissa.

# Acknowledgements

The study conducted for this dissertation took place during the most demanding times of my life so far. Therefore, the challenges of becoming a doctorate have not been the sole focus or the sole final goal of my everyday life. And for this I am grateful. Although I first did not know where to start, then what to focus on, and finally how to pack the whole thing up, it was not detrimental. Both my priority and privilege were to take care of my family and be there for my two beloved children.

Within the last few years, while I learned how to conduct research, my children learned how to ride a bike, how to brush their teeth, and do their homework. While I was nervous of my presentations, anxious of going to a conference without any familiar faces, the only possibility being to rely on myself and the good will of my colleagues, my children wrote plays and acted them out, started to walk to school and learned to trust themselves and their peers. While I learned (slowly) how to be and take part in discussions within the Academia, my children began to learn how to live, how to express their thoughts, and to take part in the world.

During the last five years, both my academic and personal selves have gotten support and encouragements and gained scaffolding relationships both from institutions and from wonderful colleagues. First and foremost, I am grateful for all the support from my supervisors Professor Anna-Maija Korpijaakko-Huuhka and Professor Sari Kunnari. Their insightful advices have helped me through the toughest puzzles. Even though I have yet to learn their pragmatic way of teaching and conducting research. Besides the support from my supervisors, I am thankful to Dean Risto Kunelius for the institutional support, which made it possible for me to explore novel technology and to take part in new inspiring ways of conducting research. In addition, I am grateful for the financial and educational support I received from the Doctoral Programme of LANGNET, to Alex Cristia for the opportunity to join the DARCLE community, and to all fellow doctoral researchers, who have broadened my understanding of science.

In this work, I had the opportunity to be the first to study and apply an interesting technology with the Finnish language and with twins. The data gathered for this research is unique; no similar data of twins exists. For this, I thank all the lovely families who trusted me enough to open their homes for me and who kindly stayed with me throughout the whole study period. I was always looking forward to the next meeting to see how the children had grown and what they had learnt. I continue to be amazed of all the love that can be heard from the recordings!

The new way of acquiring data and the extensive amount of data brought challenges that were present when analyzing and reporting the results. While acknowledging the demands of the task, I am fortunate to have had excellent advice during the work. I am deeply thankful to Dr. Melanie Soderstrom and Dr. Kaisa Launonen for their kind advice that helped me finalize the manuscript. In addition, I owe big thanks to Ph.D. Päivi Laukkanen-Nevala for the opportunity to ponder on the best ways to evaluate LENA. Your mentoring turned out to be a key factor in my learning, which took place during one of the highlight moments of the whole work process. Additionally, I thank Ansa Lilja and Jyrki Ollikainen for consultation during the statistical analyses and the LENA Foundation crew for all the technical support.

With the five years of this study, my lovely friends and colleagues have all helped me relentlessly in my journey. I thank Leena for the opportunity to focus on learning and Tarja, Anna, and Sonja for all the listening, laughs, and love – I could not have imagined getting through the last few years without your support. I am most grateful to my co-warrior and a wise colleague Katja, whom I can rely both in my worst and in my best of times. I thank Liisa and Tiina for sharing interest and for all the lovely and sometimes even heartfelt discussions - I have learned a lot alongside you. I am also thankful to my previous co-students and current clinical colleagues, who all influenced my way of thinking and being.

Lastly, I want to thank my family. Kiitos äiti ja isä, teidän ansiostanne olen kasvanut uteliaaksi ja kykyihini luottavaksi. Tukenne avulla olen selviytynyt vaikeistakin ajoista, ja ilman teitä ei tätä työtäkään olisi. Kiitos veljelleni Ramille kaikesta lapsuuden härnäämisestä; paksusta nahasta on ollut sittemmin hyötyä. Kiitos myös viime vuosien kuuntelemisesta, keskusteluista ja ihmettelyistä, joiden avulla työni parani suuresti. Kiitos miehelleni Jussille kaikista näistä vuosista, joiden aikana olet huolehtinut lapsistamme ja arkisista velvoitteistamme, jotta olen voinut uppoutua työhöni. Olet sietänyt venyviä työpäiviäni ja työmatkojani, väsyneitä kiukunpuuskiani ja epävarmuuttani äärimmäisen pitkämielisesti. Kiitos, että olet antanut minulle tärkeimpäni.

Iiris ja Tiitus, kiitos että teitte minusta äidin. Kiitos, että olette ja kasvatte. Tämä työ on teidän ansiotanne ja teidän vuoksenne valmistunut.

Tampereella, 20.10.2016

# Key abbreviations

| | |
|---|---|
| ADS | Adult-directed speech |
| AE | American English |
| AGA | Appropriate (weight) for gestational age |
| ASD | Autism spectrum disorder |
| AWC | Adult word count |
| CDEV | Checklist for Vocal Development |
| CDS | Child-directed speech |
| CHILDES | Child Language Data Exchange System |
| CVC | Child vocalizations count |
| CVM | Checklist for vocal and motor development |
| CTC | Conversational turns count |
| DLP | Digital Language Processor (LENA™ recorder) |
| DZ | Dizygotic |
| ELBW | Extremely low birth weight |
| FT | Full-term child (born 37+0 – 41+6 weeks of gestation) |
| LBW | Low birth weight |
| LENA | Language Environment Analysis (LENA™) |
| MZ | Monozygotic |
| MB-CDI | The Mac-Arthur-Bates Communicative Development Inventories |
| MLU | Mean length of utterance |
| MSL | Maximum sentence length |
| SES | Socioeconomic status |
| SGA | Small for gestational age |
| TD | Typically developing child |
| VLBW | Very low birth weight |

# List of figures

# List of tables

# Contents

# 1 Introduction

This dissertation study is one of many studies that inspect language acquisition of a certain special group of children – in this case twins – through developmental milestones and selected follow-up points, and compare the results from the measurements with information about what is presented as typical. This traditional view seeks answers to questions related to early language skills and the onset of a (possible) delay in the language development of twins. Some researchers have stated that when pre- and perinatal environmental hazards are controlled, twins' language is not delayed (Lung, Shu, Chiang & Lin, 2009), while others state that the language delay in twins is largely due to (social) environmental factors (see Thorpe, 2006, for a review). However, it is also suggested that the development of singletons and twins cannot be directly compared, as they acquire skills in profoundly different environments, and that the triadic situation familiar to twins (instead of dyadic) should be taken into consideration (Savic, 1980; Treblay-Leveau, Leclerc & Nadel, 1999; Rendle-Short, Skelt & Bramley, 2015).

Therefore, this work also aims to describe certain aspects of the language environment in twin families and to explore possible connections between natural language environment and twins' early language skills. The work has been longitudinal in nature, as it involved a follow-up of development and monthly measurements of the quantity of heard and produced speech and speech-like expressions in everyday family interaction. For this, a novel large-scale automatic method was applied. Therefore (and differently from typical logopedic research tradition), this study includes both an evaluation of the automatic method and an evaluation of the uniformity of automatic and traditional methods, which are implemented in child language development studies within academia and clinical settings. Thus, the questions related to the ontology of language acquisition and the epistemic and methodological conceptions of knowing and acquiring knowledge from phenomena related to the acquisition process are all essential.

When conducting research on child language acquisition, researchers inevitably plan, conduct and interpret results from their study with implicit or explicit conceptions of how language comes to children (Ambridge & Lieven, 2011). These ontological conceptions fall within one or both of the two major theoretical approaches questioning the classic nature-nurture problem. The first approach (generativist – nativist – universal grammar approach) sees language acquisition as an innate process, whilst the other approach (constructivist – emergentist – socio-pragmatic – functionalist – usage-based approach) highlights the role of the environment as a provider of input and scaffolding structures to support the developing child (for more detailed presentation of contrasts, see Ambridge & Lieven, 2011).

In the current study, the nativist view of innateness is acknowledged from the viewpoint of a developing physical body, genetics, and plasticity within the neural system – a device for perception and learning. However, the role of the socio-constructivist view is emphasized in regards to understanding the family environment as a learning environment, in which, on one hand, a language-acquiring child is a receiver of actions, sensations, and experiences from the environment, but is also a participant in an active reciprocal and dynamic processes of interaction. In this study, both aspects (the received input and the actions of an individual child) are seen to work as carvers of the developing neural network, and thus, have an effect on the behavior and emerging social-cognitive-linguistic abilities of the child.

Results from such learning have been traditionally measured through direct professional or indirect parental observations, which may also include an emphasis, although often implicit, on how scientifically meaningful information can be gathered, and who is seen to be able to act as an expert or provider of such information in regards to the studied subject. The conceptions related to acquiring scientific information are well related to the selection of measures, which should be able to grasp the true nature of the studied language ability. In child language studies – whether conducted in clinical practices or within academic research – this selection is often done by either choosing standardized test protocols or using different semi-structured methods, such as parent-filled questionnaires and language sample analyses, which can never truly escape the influence and interpretation of the observer. However, automatic analyses could offer a new way of dealing with the problem of biased information by allowing us to study coherence of information, and thus, serve to provide more truth-like information.

The naturalistic combination of nativist-constructivist views is in accordance with propositions from a range of laboratory learning experiments with restricted stimuli, which suggest that children segment and learn language by probabilistic pattern detection and statistical learning (e.g. Kuhl, 2004; Teinonen, 2009; Teinonen & Huotilainen, 2012). But (as articulated by Alex Cristia, 2015), child language acquisition emerges "in the wild" and as a result of everyday interaction, which is often complex in nature. Children are, for example, surrounded by multi-participant conversations and overlapping, uncoordinated stimuli, and yet, they learn to master the language they are exposed to. This poses empirical and rational challenges that need to be taken into consideration when acquiring information; how can the phenomena of child language acquisition be studied from the viewpoint of statistical learning in a natural environment?

Until recently, the field that studies children's language acquisition and natural language environments has faced limitations, which have guided researchers to either conduct more hermeneutic, qualitative research on interaction or strictly experimental investigations. Research in the field has faced several methodological and technical constraints, such as small qualitative samples and varying sample sizes for various measures, time-consuming transcription, and observations from laboratory environments, instead of natural living environments (Tomasello & Stahl, 2004; Oller *et al.*, 2010, Molemans, van den Berg, van Severen & Gillis, 2012). These methodological constraints have kept the field from pursuing a comprehensive picture of environmental effects on language development, on which Hoff (2006) states, "systematic comparative studies of children's experiences in different environments" are required, as well as valid measures to do this comparison. Until this day, we have lacked the proper means to conduct research, which would enable us to study the natural environment in which a child is acquiring language without the presence of research staff, or it has demanded an exceptional dedication to science - like living for three years in a fully wired home, which was the case in the Human Speechome Project (Roy, Frank & Roy, 2012). For this, an automated audio analysis software LENA™ and a digital recording device (digital language processor, DLP) (Warren & Gilkerson, 2008) might offer a solution.

The use of LENA has increased rapidly after its release in 2008, and it has been seen to be on the frontier of digitalization within the field of speech and language therapy, especially in the US. However, the research on reliability and validity of LENA has been fairly limited, and in most cases, such studies have been conducted with representatives from the provider, the LENA Foundation. Therefore, concerns and critique of the use and usability of LENA still exist. Fairly little is actually known about the reliability of the system and its suitability to different languages and cultures. Therefore, the explorations of its use are currently needed, as discussions about its suitability to scientific and clinical settings is emerging within the community of child language researchers.

# 2 Literature review

Twin children's systematic language studies began in the 1930's, although some records of case studies have been presented already in the 18th century (Day, 1932 a&b). A body of literature has since been published from the viewpoint of twin-singleton differences (e.g. Conway, Lytton & Pysh, 1980; Day, 1932a; Davis, 1937; Lung *et al.*, 2009; McMahon, Stassi & Dodd, 1998) and genetics (e.g. Haworth, Kovas, Harlaar, Hayiou-Thomas, Petrill, Dale & Plomin, 2009; Kovas, Hayiou-Thomas, Oliver, Dale, Bishop & Plomin, 2005; Van Hulle, Goldsmith & Lemery, 2004).

In Finland, twin studies have been largely related to the inspections on mental and physical health and conducted with two significant longitudinal cohort studies in the universities of Helsinki and Oulu (e.g. Kaprio, 2006; Trias, Ebeling, Penninkilampi-Kerola & Moilanen, 2010). However, Finnish twins' speech and language studies have been few in number. The author is aware of three case studies that have been published about the development of twins' language. One longitudinal descriptive diary study of the language development of twin boys (Räisänen, 1975), one study of the prelexical development of a twin pair (Elo & Korpijaakko-Huuhka, 2011), and one focusing on describing the acquisition of three-syllable words of a twin pair acquiring Finnish (Savinainen-Makkonen, 2000). In addition to the case studies, several master's theses (e.g. Lehtinen, 2014; Petäjistö, 2016) and four twin's language-related group studies are known; Launonen (1987) conducted a group comparison between singleton and twin children in a study for her Master's Degree, which focused on children's psycholinguistic abilities. Keinänen (2010) studied the acoustic properties of speech of monozygotic and dizygotic adult twins. Rautakoski, Hannus, Simberg, Sandnabba, and Santtila (2012) explored the genetic basis of stuttering and Latvala, Rose, Pulkkinen, Dick, and Kaprio (2014) focused their retrospective study on the onset of first words and the relationship between social behavior and teenage use of alcohol.

The present work has three main themes: special features in the language development of twins, home as a natural language learning environment, and the methods of investigating the two. Thus, the first part of this review will focus on medical factors, gene-environment discussion, and the possible influencing factors, which may have a negative impact on twin children's language development. The second part of the review focuses on language acquisition through socialization and discusses the importance of input and family environment to the language-acquiring child. In the second part, a special emphasis is given to the role of twins and older siblings, as they fundamentally influence family interaction when compared with the family dynamics of families with one first-born child (Brody, 2004). Finally, in the third part, a review of past and current research methods in child language and language environment studies is presented as well as current trends of the on-going digitalization in the field.

## 2.1 Twinship - a risk for language development?

Twin babies share at least 50 percent of their genes, their everyday home environment, and the attention and care provided by family members during the early years. Therefore, twin studies have been a popular way of conducting research, which aims to explain the environmental factors and heredity of certain traits or characteristics of human behavior. In this sense, language development and language impairment studies are not an exception. Some researchers have emphasized the role of heredity in language proficiency (Stormswold, 2001), others have highlighted the role of the environment (Thorpe, 2006, review), while still others emphasized the view, where individual trajectories in language development emerge as a result of genes, environment and experiences (Plomin, 2011; see also Plomin & Daniels, 1987). In population-based studies, low birth weight, 5-minute Apgar score, male gender, low parent education, and socioeconomic status have all been identified as risk factors for language impairments and delays, suggesting that both biomedical and social factors play a role in developmental problems (Stanton-Chapman, Chapman, Bainbridge & Scott, 2002; Korpilahti *et al.*, 2016, Wallace *et al.*, 2015). In addition, although there are differing views on how much environmental factors influence the language competence of individuals, it is an indisputable fact that language does not develop in isolation without any model from the language environment (Hoff, 2006). Thus, when the relatedness of twinship and language is questioned, the issues of heredity and biological and social environment (and the interactions and the overlapping of heredity and environment) are all essential. In the following subchapters, closer attention is paid to the gene-environment-debate, pre- and perinatal medical factors, and the implications of their effects on twins' language development.

### 2.1.1 The role and the relations of genes and the environment in twin studies

The exceptional possibility of gaining information about heritability by conducting studies with twins was first noted by Sir Francis Galton in the late 18th century (Winerman, 2004). The classical setting in twin studies relied on the notion that "identical" monozygotic (MZ) twins share all of their genes, while "non-identical" dizygotic (DZ) twins share only 50% of the genes. Thus, the comparison between MZ and DZ twins was thought to reveal whether a trait was of genetic origin or the result of environmental factors. These studies relied on the equal environment assumption (see e.g. Felson, 2014), i.e., the idea that shared factors included all the things that were common to children (e.g. the pretermity in twins, growing up in same family, attending to same school, etc.), and non-shared factors were nearly discarded or simply seen as things that are completely related to only one of the twins. In more recent studies, shared environment has been defined as any environmental factor that makes subjects similar and non-sharing any trait, which makes MZ twins different (Van Hulle *et al.*, 2004). In addition, it has also been suggested that the majority of other than genetic-based variations would be due to non-shared factors (Plomin, 2011).

Although current researchers do include both shared and non-shared environmental variables in their research designs, according to MacCoby (2006:26, review), there are still ongoing discussions about questions related to the acquisition of representative data, the definitions of shared and non-shared environments, and the interpretation of study results. In choosing the way results are presented, either the role of genes or the role of the environment is emphasized. For example, an extensive meta-analysis from twin and adoption studies concluded that heritability would explain a significant proportion (⅓to ½) of variance on the linguistic abilities for typically developing twins (Stromswold, 2001), but, in another study, it was concluded that the majority of variance in twins' language development was due to shared environmental factors (explaining 54–78% of variance) (Van Hulle et a., 2004). However, in a closer comparison, the results of Stormswold (2001) and VanHulle *et al.* (2004) are actually very much aligned. In fact, Stormswold (2006) herself has later reformulated that although genetic factors played an important role in the studies of her meta-analysis (2001), flipping of the way results from Stormswold (2001) paints a different picture by highlighting the remaining ½ to ⅔of variance not explained by genetic factors.

Besides the way of articulating results, confusion in twin studies may arise, for example, from difficulties in defining and teasing apart genes versus a shared environment and a shared versus a non-shared environment. One example of the overlap between genes and a shared environment is the case of socioeconomical status (SES). Several ways of operationalizing and measuring SES have been proposed and most typically, SES has included some quantification of family income, parental education, and occupational status (Bradley & Corwyn, 2002). SES is important, because it has been associated with maternal volubility and responsiveness (e.g. Vanormelingen & Gillis, 2016), and it has been shown to affect children at multiple levels. In addition, its effects are moderated by child and family characteristics as well as external support systems (for the effect of SES in twin studies, see e.g. Mogford-Bevan, 1999; Thorpe, Rutter & Greenwood, 2003). SES is at least partially culturally related; the relations of

SES and culture have been shown, for example, in PISA[1] studies, where country-related differences in the magnitude of SES effects on student performance scores varied between countries. SES has been reported to be significantly lower in explaining student achievement variance in, for example, Finland and Canada, when compared to US (Laurie, 2009). However, other research has also demonstrated SES to be at least partially gene-related, e.g. in IQ and education, and thus, SES can be seen to be a mix of genetic and shared environment factors (Rowe, Vesterdal & Rodgers, 1998).

The discussion on (re)defining the shared and non-shared environment has risen hand in hand with our understanding of epigenetics. For example, Plomin (2011) has argued that most of the environmental variance is actually of the non-shared variety, because non-shared environmental factors include the individual experiences of occurred events. An event or factor from the environment can be interpreted as a shared environmental effect (e.g. having the mother as the primary caretaker), but they can also involve the differentiating experiences of an individual (e.g. twins' experience of the mother's preference towards one twin over another; see Minde, Corter, Goldberg & Jeffers, 1990), which can be an important source to non-shared experience. The view of the importance of individual experience by Plomin (2011) is in accordance with the view of MacCoby (2006: 26, review), in which she concludes that "comparisons of heritability estimates based on observational reports of mother–child interaction are almost always lower than such estimates based on parent report or child report, so that observational data allow more room for shared and unshared environmental effects to be shown".

The discovery of epigenetics has been interpreted as a missing link between nature and nurture (Tammen, Friso & Choi, 2013). As epigenetic patterns may change throughout one's lifespan from early life experiences and environmental exposure, epigenetics is without a doubt one of the hot topics in behavioral sciences. Importantly, the role of epigenetics is very much in accordance with the dynamic systems and ecological perspectives on human development and interventions (see e.g. Thelen, 2005; Sameroff & Fiese, 2000), as epigenetics offer an intriguing way of measuring the effects of experiences and interventions, thus making way to evidence-based preventive interventions in behavioral sciences (Leve, Harold, Neidheiser & Patterson, 2010). However, no epigenetic transformation is possible without the mere existence of human biology or without (biophysical or social) environmental influences. Thus, the questions related to physical and social environmental factors are discussed in the following chapters.

[1] PISA = Programme for International Student Assessment (see https://www.oecd.org/pisa/aboutpisa/)

### 2.1.2 Shared and non-shared medical factors affecting the development of twin children

Twin pregnancies have elevated risks of complications both for the mother and for one or both children during pregnancy and during delivery. From all deliveries in Finland in the year 2014, 1.4% were multiple deliveries, of which 1526 children were born alive (SVT, 2015). In the cohort study of Finnish twin pregnancies, it was found that 63% of twin pregnancies included complications that required either more intensive follow-ups or medical treatment (Purho, Nuutila & Heikinheimo, 2008).

The risks of twin pregnancy include pretermity, prematurity and low birth weight, pre-eclampsia, pregnancy diabetes, maternal toxemia, pregnancy hepathosis and fetal growth restrictions (for a thorough review, see Stromswold, 2006). Out of all twin pregnancies in Finland, 40–50% of children are born as preterms (before 37 weeks of pregnancy) and by cesarean section, while on the population level, 5.9% are born as preterms and 16% via cesarean section (Uotila *et al.*, 2011; Purho *et al.*, 2008; Tiitinen, 2011; SVT, 2015). In addition to risks during pregnancy, multiple births also have elevated risks, which are evident, for example, in higher death rates for the later born B-children and, when born through vaginal delivery, later born B-twins are reported to suffer from hypoxia and complications more often than in cesarean sections. (See Purho *et al.*, 2008; Smith, Fleming & White, 2007). One in ten sections of twins are reported to be emergency sections, while only one in a hundred births are emergency sections on the population level (SVT, 2015).

**Table 1.** Definitions of categories for pretermity (WHO, 2015; Shapiro-Mendoza & Lackritz, 2012).

| Level of pretermity | Weeks of pregnancy |
|---|---|
| Full-term (FT) | 37-41 |
| Late preterm | 34<37 |
| Moderate preterm | 32<34 |
| Early preterm | 28<32 |
| Extremely preterm | <28 |

Definitions for the severity of pretermity are presented in Table 1. Although the majority of preterm twins are born as late preterms, early preterm births are also more common for twins, when compared with singletons. Purho and colleagues (2008) reported that 6.9% of Finnish twins are born as early preterms (<32 weeks of pregnancy), when on the population level, 0,8% of all children born in 2014 were early preterms. In addition, besides pretermity, low birth weight (LBW, <2500g) is also common in twins. In Finland, nearly half of the twins (40 – 42.8%) are reported to be born LBW (Stakes, 2009; Tiitinen, 2011), when on population level, 4.3% of all children are born LBW (SVT, 2015). Of all born twins, 5.3–9% have been reported to be very low weight (VLBW, <1500g), while in the year 2014, 0.7% of all newborns were VLBW (Purho *et al.*, 2008; Tiitinen, 2011; SVT, 2015).

Although preterm babies are often low in weight, both preterm and full-term children can be born small (SGA, birth weight <10 percentile), appropriate (AGA) or large for gestational age (LGA>90th percentile). SGA is caused by intrauterine growth retardation (IUGR), which may have origins in medical conditions of the mother, placentae, or physiological assets of the unborn children (Sharma, Shastri & Sharma, 2016). IUGR twins have higher risks, for example, for respiratory distress syndrome and intraventricular hemorrhage (Yinon, Mazkereth, Rosentzweig, Jarus-Hakak, Schiff & Simchen, 2005), and SGA children have a significantly increased risk of hypoglycemia (De Bruin, van der Lugt, Visser, Oostdjik, van Zwet, te Pas & Lopriore, 2015).

SGA children are reported to be at an increased risk of poorer communication skills (Partanen *et al.*, 2016), and all of the previously mentioned conditions affect twins more often than singletons. For example, mothers of twins have reported to suffer from pre-eclampsia more often than mothers of singletons (Purho et al 2008; Luoto, Kinnunen, Koponen, Kaaja, Männistö & Vartiainen, 2004); early stage pre-eclampsia in particular can affect histopathological formation of placentae, cause placental ischemia, and thus, have an effect on the growing fetuses (Karikoski, 2011). With twins in monochorionic diamniotic pregnancy, placental problems can also cause twin-to-twin transfusion syndrome (TTTS), which can further cause mild to severe growth problems.

As presented above, twins are faced with several medical risks, which may lead to developmental problems, including difficulties and delays in communication, speech, language (see e.g. Stanton-Chapman *et al.*, 2002, for low Apgar scores; Bishop, 1997, for medical risks and twin's language), and mother-child interaction (Korja, Latva & Lehtonen, 2011; Muller-Nix, Forcada-Guex, Pierrehumbert, Jaunin, Borghini & Ansermet, 2004; Schermann-Eizirik, Hagekull, Bohlin, Persson & Sedin, 1997). Additionally, as twins are often overrepresented in preterm groups (Foster-Cohen, Edgin, Champion & Woodward, 2007), the question of the relation between medical and other risk factors and twins' later development is essential. From previous studies, it is already known that children of low birth weight have an increased risk of medical conditions in pre- and neonatal stages, when compared with children of normal birth weight (e.g. Rutter, Thorpe, Greenwood, Nothstone & Golding, 2003; Stromswold, 2006, review); the relations between gestational age (GA) and/or birth weight are also established in many studies in several social-cognitive-linguistic domains.

Apart from typically developing children, the vast majority of research in the field of language acquisition has focused on very and extremely preterm and VLBW and ELBW children, although late preterm infants are also reported to be at risk of unfavorable developmental outcomes (McGowan, Alderdica, Holmes & Johnston, 2011, review). Finnish preterm ELBW children have been reported to begin canonical and variegated babbling similarly as FT children, but to produce their first words later (M=13 months) than FT children (M=11 months) (Törölä, Lehtihalmes, Heikkinen, Olsèn & Yliherva, 2012b). In addition, preterm children have been reported to show lower social responsiveness (De Schuymer, De Groote, Beyers, Striano & Roeyers, 2011) and very and extremely preterm children's receptive and expressive lexicons is shown to develop later than those of FT children (Foster-Cohen *et al.*, 2007; Kern & Gayraud, 2007; Vohr, Garcia, Coll & Oh, 1988). Preterm children are also reported to be less skilled in the use of words and word endings (Foster-Cohen *et al.*, 2007) and have less complex sentences than FT children (Foster-Cohen *et al.*, 2007; Kern & Gayraud, 2007; Kunnari, Yliherva, Paavola & Peltoniemi, 2012). Additionally, SGA children have been shown to have more linguistic and motor problems at school age (Yliherva, Olsèn, Mäki-Torkko, Koiranen & Järvelin, 2001).

### 2.1.3 Twin children's language development

Although twins are faced with several health risks early on, not all twins suffer from such disadvantages. Previous research has suggested that if there are no major complications during pregnancy, the twin situation itself seems to be meaningless to children's early development (Tomasello, Mannle & Kruger, 1986; Lung *et al.*, 2009) or that birth weight and pretermity would be at least the best explanatory factors of developmental delays and problems in twins (Anand, Platt & Pharoah, 2007). However, it should also be noted that although present, health risks do not always affect language acquisition in twins. For example, no relationship between twins' language abilities and pretermity was found in the studies of Conway, Lytton, and Pysh (1980) and Stafford (1987). Additionally, the studies of Mittler (1970) and Bishop (1997) found no to little relation with children's language abilities and complications in twin pregnancy, delivery, or Apgar scores. Although Apgar-scores were not found to explain later development, Bishop (1997) found a close to significant difference of exposure to maternal toxemia during pregnancy between children grouped as having specific language impairment (SLI) and children grouped as typically developing (TD) controls. In addition, Mittler (1970) discovered that psycholinguistic scores were lower for preterm children, when compared with full-term (FT) children.

However, besides twins with medical risks, healthy twins also have been suggested to have a delayed language development when compared with singletons (Rutter *et al.* 2003, Thorpe, 2006, review). There is also evidence from a follow-up study, which suggests that the lag in twins' development, and especially in phonological development, would continue to manifest in poorer literacy skills in the early school age (McMahon *et al.*, 1998). In fact, the late language emergence (Rice, Zubrick, Taylor, Gayan & Bontiempo, 2014; Thorpe, 2006, review) and the disturbances in phonological development have been suggested to be typical features in twin children's distorted language development (Hua & Dodd, 2000; McMahon *et al.*, 1998). Previously, a phenomenon nowadays considered to represent deviant phonological development in twins was formerly thought to be related to the growth environment, where twins acquiring language would develop a "secret twin language" called *cryptoglossia or cryptophasia* (Bishop & Bishop, 1998 Hua & Dodd, 2000; Luria & Yudovitch, 1959; McEvoy & Dodd, 1992; Rutter *et al.*, 2003).

It has been suggested that twins with normal language development would come from middle-SES families with high parental education (Mogford-Bevan, 1999). However, it is acknowledged that more information is needed on the possible long-term effects delayed language development might have up to adolescence and adulthood (Thrope, 2006, review). But, beside the later outcomes, there is also a lack of information about the earliest stages of twin children's language development, although some work has been performed on toddler-aged twins, and a great body of work has been presented on the language development of preschool-aged children. These studies have had contradicting findings on the language development of twins. Some have found delays, while others have not, and some have also reported atypicalities in twin children's development. Rutter and colleagues (2003) reported a mild delay (1.7 months) in the language development of late preterm and full -term twins with no medical conditions at the age of 20 months with an increasing lag of 3 months by the age of 3 years. A similar lag in development has been found earlier by Stafford (1987). Stafford conducted a study with twins and singletons ages 24–36 months and concluded that twins were 2.5 months behind singleton controls in comprehension and 3.7 months behind in expressive development. But, as Stafford points out, twins' scores still remained within normal variation. In addition, Kobayashi, Hayakawa, Hattori, Ito, Kato, Hayashi, and Mikami (2006) did not find developmental delays in their study conducted using an ITPA psycholinguistic test with three- and four-year-old Japanese twins, as well as Garitte, Almodovar, Benjamin, and Canhao (2002) in their study conducted with four- and five-year-old French twins.

However, although information about twin-singleton differences in later stages exists, there is a lack of information about the onset of the delay, as the prelexical stages of twin children's language development have also been rarely studied. This information would be of great importance, as the origin of language development can be traced back to infancy. In addition, most of the prelexical studies are comparative inter-twin case studies, such as the ones that have been conducted on the prelexical vocal development of Finnish twins (Elo & Korpijaakko-Huuhka, 2011), twins living in a bilingual environment (Zlatic, MacNeilage, Matyear, & Davis, 1997; Mayr, Price & Mennen, 2012), and twins differing in auditory function (Kent, Osberger, Netsell & Goldschmidth Hustedde, 1987).

For this review, only one group study was found on the prelexical development of twins, and the onset of first words was inspected in two studies. Nan, Piek, Werner, Mellers, Krone, Barret, and Zeekers (2013) conducted a questionnaire study with developmental follow-up of up to 24 months of age. They found a delay in early communication skills at the ages of three months, but twins were found to catch up the delay already before the age of 6 months. However, Mittler (1970) found the first words of twins to be delayed. And recently, a Finnish retrospective questionnaire study found that parents of twins reported their children's first words to appear at the mean age of 14.6 months (Latvala *et al.*, 2014), which was approximately 4.5 months behind the mean age of the onset of first words (M=11.0) previously reported (as manifested in Finnish children) (Lyytinen, Ahonen, Eklund & Lyytinen, 2000). In addition, Latvala and his colleagues (2014) discovered that twin girls spoke their first words earlier than twin boys, and that there were no constant statistically significant group differences for birth order. However, Latvala and colleagues reported that twins of lower birth weight began to produce their first words later than twins with more appropriate birth weight.

Besides inter-twin case studies, nearly all of the studies of twins' language development have used measures, which have been normed based on data from mainly singletons. In addition, almost all of the language acquisition studies of twins have been conducted in dyadic settings. It is, therefore, in question whether twins might have had a disadvantaged position in such studies, and whether development should be assessed in their natural, triadic contexts (Tremblay-Leveau *et al.*, 2009). The profoundly different social environment may guide twins' language to develop to face the everyday interactional challenges, which are of different linguistic demand than for singletons. Therefore, to understand twins' language development, the importance of social developmental factors need to be further inspected.

## 2.2 Language acquisition through socialization

As presented above, healthy twins are also at risk of delayed language development. Thus, the explanations of the delay must be sought from the everyday experiences in twins' social environments and practices present in their families, where children are socialized through and to acquire language (Kulick & Schieffelin, 2004; Tomasello, 2003; Schieffelin & Ochs, 2008).

The term "socialization" refers to processes, where naïve individuals are taught the skills, behavior patterns, values, and motivations that are needed in order to become a competent operator in the culture in which the child is growing up (MacCoby, 2006:13). The theoretical concept of socialization was introduced in the 1980's to enrich the psycholinguistic literature of language acquisition and the anthropological literature on child socialization, although socialization studies had been initiated in a collaborative project with psychologists, anthropologists, and linguists already in the 1960's (Schieffelin & Ochs, 2008). However, the basis of the theories of social learning can be traced back to the work of Vygotski (1982), Bandura (1971), and Bruner (1983).

In the family context, socialization to language and communication is present, for example, in contingent dyadic social interactions, where both the parent and the child actively respond to each other's speech (Golinkoff, Can, Soderstrom & Hirsch-Pasek, 2015, review). However, the scaffolding support to a language-acquiring child can manifest itself in different ways, especially if the child is an only child or if the child has siblings, since having an older sibling is presumed to have an effect on a child's language environment (Oshima-Takane, Goodz & Derevensky, 1996). Thus, first-born twins are not in a similar position as singleton second-borns, but it is unclear what kind of difference lies when growing up with a twin or with a twin and an older sibling.

The theoretical model of socialization has been criticized for placing the child in the role of a passive receiver, instead of being an active participant (Stewart, 2000). However, the impact of the social structure on infant learning is diverse, including the child's trust in his or her caretakers, the construction of common ground between the caretakers and the children, the facilitation of development, parental fine-tuning and scaffolding properties of child-directed speech (CDS) as part of the qualitative and quantitative factors of input (MacWhinney, 2014). In addition, as pointed out, for example, by Lytton (1980: 3), any relationship includes transactions between participants, and thus, socialization can also be viewed as a reciprocal process with active participants. Therefore, all the assets, previous experiences, and individual thoughts and acts of the child are always present in socialization processes.

To this date, the majority of research conducted on twin children's socialization consists of studies related to cultural socialization practices (e.g. Goshen-Gottstein, 1981; Stewart, 2000) and behavioral problems in adolescence (instead of the process of learning language through early socialization within the family environment). In addition, the processes of socialization in relation to the development of twin children's language has been studied from the viewpoint of twin-singleton differences by Hugh Lytton (1980). And, although the amount of literature could suggest otherwise, Lytton, Conway, and Sauve (1977) have suggested that socialization practices would become more influential to twins' development than, for example, social class or pre- and perinatal biological environment.

### 2.2.1 The importance of input quantity and quality

Children need input to be able to acquire language (Hoff, 2006), and it is suggested that children acquire language in everyday social settings by pattern detection and statistical (probabilistic) learning (Kuhl, 2004; Frank, 2012). It has been suggested that parents intuitively help their children in such processes using child directed speech (CDS) (Trainor & Desjardins, 2002; Yurovsky, Doyle & Frank, in review). CDS differs from adult-directed speech (ADS) in several ways (Hills, 2013; Soderstrom, 2007). CDS has, for example, shorter utterances, longer pauses, higher and more varied pitch, and more associative components, word repetitions, and more context-dependent speech – all assets that may help the child to identify meaningful segments from speech and process components from speech more easily than from ADS. Thus, it could be concluded that the qualitative properties of CDS could make statistical language learning easier, and this has been shown to be the case, for example, for word acquisition: 1) Word frequency, repetitions, and associative structure have been reported to better predict, for example, word acquisition, when compared with ADS (Hills, 2013), 2) the quantity of certain words within child-directed speech has been shown to predict the age of acquisition of the word by the child (Huttenlocher, Haight, Bryk, Seltzer & Lyons, 1991), and 3) input within word class frequency is shown to correlate with child language development (Goodman, Dale & Li, 2008).

Both mother and father are shown to modify their speech when talking to children, although contradictive findings do occur (see Soderstrom, 2007, for review). Parents use CDS intuitively, for example, by adjusting the complexity of input well before and after a child has acquired a word (Roy, 2009), and researchers have even been able to predict the acquisition ("birth") of a word of a child by analyzing spatio-temporal information from ultradense samples gathered in everyday linguistic settings from a family home (Roy, Frank, DeCamp, Miller & Roy, 2015). Thus, everyday repeated occasions with interaction might offer the key component to understand the social basis of language acquisition. In fact, in families where children are engaged in everyday activities, such as shared reading and family mealtimes, children are reported to have better language skills than children with less everyday family interaction (for synthesis on the subject, see Dunst, Valentine, Raab & Hamby, 2013), and children exposed to a higher-level of caregiver speech are found to do better on language tests than children with less access to caregiver speech (McCartney, 1984). Additionally, restrictions in receiving input have been found to have effects on prelinguistic development (Koopmans-van Beinum, Clement & van den Dikkenberg-Pot, 2001; Oller & Eilers, 1988), and negative associations for child language development have been reported for television exposure, decreased amount of adult words, (Christakis, Gilkerson, Frederick, Garrison, Xu, Gray & Yapanel, 2009; Hart & Risley, 1995), and conversational turns (Christakis *et al.*, 2009;

Ambrose, VanDam & Moeller, 2014). However, studies that address the meaning of input to a child from both quantitative and qualitative aspects generally imply that "the quantity of input is not the whole story" (Rowe, 2012), and the importance of both frequency and quality of input is supported, for example, by results from the famous studies of Hart and Risley (1995) and the Human Speechome-project (e.g. Roy, 2009).

When making assumptions about whether input frequency counts, a careful consideration must be paid on research designs. What is counted as input (overheard speech and/or child-directed speech), and who are accepted to act as providers of the input, i.e. is the research studying dyads, children with or without siblings, children within diverse families, or children participating in other groups. The issue of input quality and quantity was recently investigated in a study by Ramirez-Esparza, Garcia-Sierra, and Kuhl (2014). They analyzed both the quality of input (standard vs. parentese) and social interactional situations, when the child was hearing the input (dyads vs. groups with two or more adults). Ramirez-Esparza and colleagues found no significant effect of raw quantity of speech input (including overheard speech) on children's vocabularies at 24 months. Rather, they found that infants who were interacting in dyads and hearing exaggerated parentese were more productive in concurrent speech and had larger vocabularies at two years of age, when compared with children hearing standard speech or parentese speech in groups.

The results from the study by Ramirez-Esparza and colleagues (2014) are consistent with the results from a study by Weisleder and Fernald (2013), who reported that children who heard more child-directed speech in low-SES families were more efficient in processing familiar words in real time and had larger vocabularies at 24 months of age than children who heard less care-giver speech. However, Weisleder and Fernald (2013) found no associations between the amount of overheard speech and vocabulary size. Apart from the findings highlighting the importance of CDS as a facilitator to better language outcome, the evidence from Hart and Risley's (1995) study suggests that the overall high quantity of speech in family environments is related to the rich quality of speech (see also Hurtado, Marchman & Fernald, 2008). Therefore, speech quantity measures might serve as an indirect way to study the richness of language environments within family homes (see also Vanormelingen & Gillis, 2016, for input quantity and quality in families differing in SES).

### 2.2.2 Special interactional features of twins: shared and non-shared social environmental influences in twinship

Mother-child interaction has been shown to predict language development both for twins and for singletons (Thorpe *et al.*, 2003). However, the social environment in twin families has been reported to differ from families with singletons, as parents of twins are obliged to take care of the needs and demands of two children of same age and developmental levels (Mogford, 1993; Tomasello *et al.*, 1986). As shared environment has been shown to be a dominant factor in early language development in twins (Hayiou-Thomas, Dale & Plomin, 2012), understanding the influence of twinship on the mother-child interaction and on the interaction between the child and the child's immediate social environment in general is crucial.

The twin situation creates a unique interactional environment, where a carer and the twins often form a communicative triad. According to Mogford (1993: 87), the triadic twin situation could affect language development at least in two ways. Firstly, triadic twins may talk less because their closeness reduces the need for verbal development, opportunities, and interest in communicating with others. Secondly, the limited attention from parents providing care for both children causes parents not to be able to spend as much time with one child as parents of singletons (Mogford, 1993: 87). The latter hypothesis has received more support from the research community, as twins may still be more motivated in interacting with a present adult than with their co-twin (Savic, 1980:73). This may be the case even when mothers of twins have to share their attention more often with both children than mothers of singletons (Thorpe *et al.*, 2003). This may affect the behavior of mothers, but also the behaviors of children, as communicational challenges differ from dyadic situations, most commonly described in scientific literature.

Parents of twins have been reported to hold and look at their children less often and to spend less time interacting with their children, when compared with parents of singletons (Holditch-Davis, Roberts & Sandelowski, 1999). Twins have also been reported to get more experiences from interrupted conversations and less dyadic interactions with carers, which has an effect on the amount of experiences of joint attention occurring in dyadic interaction (Tomasello *et al.*, 1986). In addition, mothers of twins have been reported to be less infant-focused, less responsive, less accepting, to show less involvement with their children, and to provide a narrower range of experiences to their children at an early age than mothers of singletons (Butler, McMahon & Ungerer, 2003; Thorpe *et al.*, 2003).

Parent responsiveness enhances child vocalizations, infants' mapping of word referents, growth in vocabulary, and pragmatics (Goldstein, King & West, 2003; Gros-Louis, West & King, 2014; Tamis-LeMonda, Kuchirko & Song, 2014). However, mothers of twins have been observed to be more directive, use less infant-focused speech, and to attribute fewer questions and agency to their infants than mothers of singletons (Butler *et al.*, 2003). In addition, it has been reported that twins also get less motivation from mothers to observe toy, fewer invitations to look at picture, share a book and participate as an active member in conversations at an early age (Butler *et al.*, 2003; Thorpe *et al.*, 2003). Therefore, the triadic situations of twins' everyday life seem to be impoverished, when compared with singletons.

However, there is also evidence against the impoverishment thesis and suggestions to take the interactional abilities of twins into account when comparing twins and singletons, as twins have been reported to have acquired a different type of joint attention model and linguistic skills already before the age of two, when compared with singletons (Rendle-Short *et al.*, 2015; Tremblay-Leveau *et al.*, 1999). These skills were found to help twins join, follow, and take turns within multispeaker conversations, as well as cope with the information received from multiple participants (Savic, 1980; Tremblay-Leveau *et al.*, 1999), but such skills were not found in singletons in triadic situations. However, similar findings of pragmatic skills have been reported from second-born singleton children (Dunn & Schatz, 1989), suggesting that language proficiency develops through socialization to answer the demands of interactions in a child's everyday life. And, in fact, these skills might deserve to be evaluated in a formal assessment as "the ability to participate effectively in a multispeaker world – to join and contribute to talk between others – is a developmental achievement of considerable importance to all children" (Dunn & Shatz, 1989: 399). This ability has also been shown to relate to children's later peer and social development (Hedenbro & Rydelius, 2014).

### 2.2.3 Linguistic environment in a family with siblings: implications from twin and non-twin sibling studies

As the family home is the main environment for most of the preverbal children, it is essential to acknowledge the importance of all family members instead of focusing solely on mother-child interaction or even parent-child interaction – after all, the majority of individuals grow up in families with siblings (Brody, 2004; Dunn, 2006). In addition to the mother's, the father's engagement has been suggested to be positively associated with social, behavioral, psychological, and cognitive outcomes of children (Sarkadi, Kristiansson, Oberklaid & Bremberg, 2008, review). And, in addition to parents, siblings may have an important influence on child language development (Soderstrom, 2007, review) as they are to the socialization process of the younger ones in general (Kramer & Conger, 2009). Therefore, it is important to acknowledge that the experiences of second-born children differ from the experiences of first-born children.

Having an older sibling has an influence on both the younger sibling and the parent's behavior, which is also evident in verbal interactions between family members. The presence of an older sibling affects the frequency and quality of spoken input to the younger child (Oshima-Takane *et al.*, 1996; Woollet, 1986). Mothers have been reported to speak less and direct fewer words to younger children in a triadic situation, and older siblings are reported to speak infrequently to younger siblings, while ignoring the majority of younger sibling utterances (Oshima-Takane & Robbins, 2003; Tomasello & Mannle, 1985; Woollett, 1986). Estimates for spoken utterances from the mother to the older sibling have been reported to be from 40 to 67 percent (⅔) of all speech, whereas speech directed to the younger sibling is estimated to be only 33 percent (⅓) of all speech (Oshima-Takane *et al.*, 1996; Woollett, 1986).

Although the triadic interaction between the mother and her twins is suggested to differ from dyadic situations between the mother and children of different ages (Barton & Tomasello, 1991), similarities with non-twin-sibling studies also exist. Mothers of twins have been reported to speak less to children and to use shorter utterances (Conway *et al.*, 1980), as do mothers of second-born singletons, when compared with mothers of first-born children (Woollet, 1986). On the other hand, other studies have found mothers of twins to speak as much as mothers of singletons, but the amount of overheard speech has been greater (Tomasello *et al.*, 1986; Shneidman, Arroyo, Levine & Goldin-Meadow, 2013). No significant differences in the mother's mean length of utterance (MLU) was found in a study between mothers of twins and singletons (Tomasello *et al.*, 1986), but mothers of twins have been reported to use less complex utterances than mothers of singletons (Conway *et al.*, 1980). In addition, qualitative differences in the language environment for second-born children has been suggested to result in different language skills compared with first-borns. In a sibling context, younger children, for example, hear themselves being referred to as a third partner (Woollet, 1986), and children with older siblings are reported to have more advanced pronoun production skills due to overheard speech (Oshima-Takane *et al.*, 1996), just as twins in triadic settings (Savic, 1980).

Even though older siblings address the majority of their speech to the present adults, there is evidence suggesting that they modify their speech while talking to their younger siblings (Woollet, 1986) and that child directed speech (CDS) from siblings would be beneficial to children's development (Shneidman et. al., 2013). For example, older siblings shorten their utterances when they speak to younger siblings, and both mothers and older siblings have been reported to direct more social-regulative language in triadic situations to young siblings, whereas older siblings directed more metalinguistic and referential language to their mothers (Oshima-Takane & Robbins, 2003). In addition, first-born infants are reported to imitate their parents, while infants with siblings imitate sibling and adult-sibling behavior (Barr & Hayne, 2010). And, although younger siblings have been found to talk less in the presence of an older sibling, second-born children are shown to have better conversational skills when compared with first-borns (Hoff, 2006, review; Woollet, 1986) – just as twins were more skilled in multiparticipant interaction, as discussed in chapter 2.1.3.

## 2.3 How to study early child language acquisition and everyday environments as learning environments

After the falsification of the discontinuity hypothesis[2] by Jakobson (1968), a broad interest in prelexical development and its usability in detecting atypical development emerged. From these studies, it is already known that babbling has several features, which resemble speech characteristics, and that there are several universal similarities in babbling (e.g. Davis, McNeilage, Matyear & Powell, 2000; Rothgänger, 2003), but also language specific elements (e.g. Lee, Davis & MacNeilage, 2009). In addition, the onset of babbling (Oller, Eilers, Neal & Schwartz, 1999), the rate of vocalizations (McCathren, Yoder & Warren, 1999), and the number of consonants in phonetic inventories in babbling are shown to be related to later vocabulary (Stoel-Gammon, 1991), as are syllable structures for expressive language (Paul & Jennings, 1992). It is also known, that there is a causal relationship between early language skills and later linguistic and cognitive development – affecting, for example, school readiness (e.g. Forget-Dubois *et al.*, 2009; Roulstone, Law, Rush, Clegg & Peters, 2011).

In communication disorders, a primary goal of research should be in the development of causal explanations of disordered behaviors we observe (Duffy, Watts & Duffy, 1981). Additionally, research is needed to identify what actually accounts for disordered behavior, if the interventions we offer to recognized disorders are effective, and whether the interventions we implement have sustainable effects on the reduced severity of the effects of the disorder. These requirements are also true for the studies on the early stages of child development, which are often motivated by the desire to predict later developmental outcomes, and work on behalf of early interventions, which may be attenuated, if the need is identified.

There are three broad approaches to the assessment of language in young children, which are used by speech and language therapists. Analyses of language samples, parent/carer reports of language performance or concern about language skills, and standardized norm-referenced measures (Law & Roy, 2008). In addition to these clinically used approaches, child language acquisition researchers use a broad range of experimental methods, which are not in the repertoire of clinical practitioners. However, the possibilities of measuring the effects of ecological interventions are few in number, although "all theories of language acquisition acknowledge the necessity of input" (Ely & Gleason (1995), and the field of speech and language therapy is more and more about programming services into the immediate environment surrounding the children with slow or disordered speech and language acquisition (Pickstone, Goldbart, Marshall, Rees & Roulstone, 2009).

---

[2] Jakobson (1968) suggested that babbling was not related to later language development

### 2.3.1 Observational and qualitative studies

Child language studies are often conducted with contemporary data and are therefore inevitably influenced by the current cultural environment and tools available at the time of the study. According to Ingram (1989: 7), the field has had three major periods, each being defined by certain methodological similarities within conducted studies; a time of diary studies from the late 18th century until the onset of 19th century (when research of child language studies originated), 2) a time of large sample studies in the middle of 19th century, and 3) a time of longitudinal studies from the 60’s to then present time (which was the time of Ingram’s book, i.e. 1989).

As the first diary studies of child language emerged, so did the first study of twins’ language acquisition. Horatio Hale described one twin pair’s distorted speech development in the year 1886 (Day, 1932a). The descriptive nature was, in fact, characteristic to child language studies at that time, which were to have only little concern over theoretical construction or emphasis on the environmental factors (Ingram, 1989: 8–10). According to Snow (2014: 117), the nowadays well-recognized scaffolder – the linguistic environment (namely, child-directed speech) –in the mid 70’s was only “a modest body of work documenting that speech addressed to young children was generally grammatically simple and lexically redundant,” but that “no one had actually demonstrated that these adaptations made any difference”. In addition, the environmental effects were set aside because the research community failed to make a distinction between environmental and genetic effects on child development (Snow, 2014). This, however, was not unique to child language studies, but similar phenomena have been identified in the field of socialization studies, also focusing on other aspects apart from language acquisition (MacCoby, 2006: 26).

Although diary studies had substantial influence as the originators of the field, early studies suffered from several disadvantages; entries were based on memory and interpretations of the observer, the annotation of prosody and phonology was imprecise, para- and extralinguistic information (e.g. gesture, gaze, repetitions, and false starts) could not be sufficiently captured, errors and analogical forms were overrepresented, while correct forms were underrepresented, and the samples were small (Slobin, 2014). Some of these methodological problems were solved in the late 1950’s, when researchers with qualitative study designs began to pay more attention to systematic data collection and planning of follow-ups (Ingram, 1989: 12). In addition, the field brought the inter-child variation into discussion and went on to conduct in-depth linguistic analyses on child language data (Ingram, 1989; Fletcher, 2014).

After the introduction of tape-recorders, it became possible to conduct language sample analysis (LSA) on transcribed data (Behrens, 2008). However, even to this date, researchers and clinicians seem to conduct language sampling and LSA in very different manners. The problem, however, has been identified in Finland as well (Saaristo-Helin & Savinainen-Makkonen, 2008); and researchers became interested in clarifying, for example, how sample size should be defined (utterances vs. minutes), what kind of sample is sufficient, and what kind of sample length is suitable for children with different developmental and medical conditions, different ages, research interests, and ways of analyzing samples (e.g. Heilmann, Nockerts, & Miller, 2010; Guo & Eisenberg, 2015; Molemans et al; 2012; van Severen, van den Berg, Molemans & Gillis, 2012).

The questions related to sampling are essential, as sample size and density have been shown to affect the reliability of results (Adolph & Robinson, 2011; Guo & Eisenberg, 2015; Molemans *et al.*, 2012). In addition, naturalistic speech analysis could be very cost-effective in the long term (Ambridge & Rowland, 2013, review), if conducted in a unified manner. When language sample analysis also includes calculations of ratios, describing language proficiency, the information can be compared between participants and studies. Such ratios include, for example, measures of prelexical development through canonical babbling ratio CBR, true canonical babbling ratio, and TCRB (see Molemans *et al.* 2012, for suggestion on reliable detection of babbling), mean babbling level MBL (Stoel-Gammon, 1989), and mean length of utterance (MLU) before (Fagan, 2009) and after words (Brown, 1973). Besides MLU and its variants (mean length of utterance morphemes, MLUm; mean length of communication units, MLCU; phonological mean length of utterance, PMLU; maximum sentence length, MSL), other widely used measures at the later stages of development include syllable structure level (SSL) (Paul & Jennings, 1992; Morris, 2010). However, all mentioned ratios still require hand-coded transcriptions, even if they are calculated using tools for computer-assisted and systematic LSA (chapter 2.3.3).

### 2.3.2 Checklist studies

When the behaviorist approach emerged in the field of child language studies in the beginning of the 19th century, researchers began to lay special emphasis on what could be seen as normal behavior; many standardized tests were developed during the period of large samples (Ingram, 1989: 12–13). During the era, researchers began to conduct comparative studies on, for example, typical, talented, and lower-class groups of children (Ingram, 1989: 14). It was at this time when the first comparative paper on twin-singleton differences was also published (Day, 1932a). From the onset of the 1920's till the end of the 1950's, researchers conducted studies with large samples and with well-planned study designs, but one weakness of these studies lied in hand-written notes, which were done due to the lack of modern devices (Ingram, 1989: 16). Thus, effective methods needed to be developed.

Although numerous checklists have been presented in scientific literature, child language surveys have been quite rare in Finland, and there are currently only a few checklists that have been validated for Finnish and are used in clinical practices. Prelinguistic checklists in Finland include the *Checklist of development of early vocalizations* (later, CDEV; (Lyytinen *et al.*, 2000) and the more recent *Esikko* – the Finnish version of the Infant–Toddler checklist by Prizant and Wetherby (Laakso, Eklund & Poikkeus 2011). In addition to developmental checklists, clinicians screen for developmental milestones, including the onset of canonical babbling, pointing, gesture communication, and first words.

Parents have been reported to be reliable in identifying the onset of canonical babbling, which begins with typically developing children by the 10th month (Oller, Eilers & Neal Cobo-Lewis, 1998). However, the relations between the onset of reduplicative babbling and later vocabularies differ between studies. Some studies have not found the onset of reduplicative babbling to correlate with the onset of word comprehension and production (Fagan, 2009; Oller, Levine, Cobo-Lewis, Eilers & Zurer Pearson, 1998), while other studies have found milestones to predict later language development (Lyytinen, Poikkeus, Leiwo, Ahonen & Lyytinen, 1996), or that the late onset of canonical babbling was related to smaller productive vocabularies at 18, 24, and 30 months (Oller *et al.*, 1999). Currently, late talkers can be identified from typically developing peers from 18–24 months onwards (Paul, Sgangle Looney & Dahm, 1991). For the earliest age stages, that is, before the age of two years, the only validated checklist for language development in Finland is The Finnish version of the MacArthur-Bates Communicative Development Inventories[3] (Lyytinen, 1999), originally published for American English in the year 1994 (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994).

#### 2.3.2.1 Description and reliability of the Checklist of Development of Early Vocalizations

The CDEV is a part of a parental checklist *Ääntelyn ja motoriikan kehityksen seurantamenetelmä - the Checklist of vocal and motor development* (later, CVM; translation and abbreviation by the author) developed during the Jyväskylä Longitudinal Study of Dyslexia (Lyytinen *et al.*, 2000). In addition to the CDEV, the CVM questionnaire has questions for gross and fine motor skill development (Lyytinen *et al.*, 2000), which are not further addressed in this study.

[3] The MacArthur-Bates Communicative Inventories have been referred to in several different abbreviations, some of which are used globally (CDI, MCDI, MB-CDI), and others highlight the language-specificity in the abbreviation of adaptation versions (e.g. FinCDI for Finnish version). This study adopts the newest version of the abbreviation and refers to the MacArthur-Bates Communicative Development Inventories with the abbreviation MB-CDI.

The CDEV is based on parent detection of the onset of developmental milestones, but it asks the time of the first three observations of a skill instead of just the first, thus building a more reliable picture of a child's acquisition of new skills. For the precursors of speech, the CDEV screens for the onset of vocalizations with a passive vocal tract within one breath group, the onset cooing and raspberries, vocal imitations, reduplicative babbling, variegated babbling, and protowords. In addition, it screens for the onset of meaningful gestural communication and communication through actions (pointing, head shaking, waving, etc.).

The reliability and validity of CDEV has been evaluated by the validation study of the method (Lyytinen *et al.*, 1996; Lyytinen *et al.*, 2000). The manual of the normative study presents statistical dispersion and values for normal variation of the appearance of vocal milestones. In addition, Lyytinen and colleagues (1996) presented the median ages and standard deviations for the emergence of vocal behaviors, from which, the onset of reduplicative (*Mdn*=27.0, *SD*=8.7) and variegated babbling (*Mdn*=32.0, *SD*=9.1) are of importance for the present study. In addition to statistical dispersion, the CDEV manual (Lyytinen *et al.*, 2000) presents correlations between the whole CDEV and later language measures. The total score of CDEV is reported to correlate with the Reynell Developmental Language Scales (RDLS, version not reported) and the mental and motor portions of Bayley Scales of Infant Development (BSDI) at 18 and 24 months (all correlations $r_s$=.25–.30, p=.05–0.1). In addition, the CDEV total score statistically significantly correlated with several parts of MB-CDI at different ages.

In the normative data, the total score of CDEV was statistically significantly correlated with the receptive vocabulary measure of the MB-CDI at 12 ($r_s$=.51, $p$<.001) and 14 months ($r_s$=.46, $p$<.001) and with expressive vocabulary measured with MB-CDI at 12, 14, 18, 24, and 30 months ($r_s$=.26-.33, $p$=.01–.05 in all). In addition, the total score of CDEV correlated with the MB-CDI measured maximum sentence length (MSL) at 18 ($r_s$=.24, $p$<.05) and 24 months ($r_s$=.26, $p$<.05), but not at 30 months. However, the total score of CDEV did correlate with the sum variable of morphological skills at 30 months ($r_s$=.30, $p$<.05), but not with the sum of morphology at 18 and 24 months.

The CDEV manual does not, however, provide information on the relations of later development and the single skills screened in the checklist. However, Stolt, Lehtonen, Haataja, and Lapinleimu (2012) did study the relations of vocal milestones and later development using the CDEV with children acquiring Finnish, of very low birth weight (VLBW) children, and children of normal birth weight (NBW). Stolt and colleagues (2012) discovered that there were statistically significant correlations between the onset of babbling and later language (measured with RDLS III). In addition, the onset of protowords was statistically significantly correlated with lexicon and maximum length of three longest utterances at the age of 2.0 for VLBW children, but not for controls. Although there were differing results for the relations of the onset of reduplicated babbling and later language proficiency, no statistically significant group differences were found between VLBW and controls for the rate of acquiring early vocalizations when analyses were conducted using corrected ages for VLBW children.

#### 2.3.2.2 Description, reliability, and validity of the MacArthur-Bates Communicative Development Inventories

The development of the MacArthur-Bates Communicative Development Inventories began in the 1970's when the original version was developed for research purposes (Fenson *et al.*, 1994). Before the MB-CDI, researchers had been mainly using limited screening instruments and time-consuming language sample analyses, as parent reports were yet untested at that time. During the twenty years of development, the MB-CDI developed from early stage free-form interviews firstly, to a structured checklist for oral administration, and then to the final version of self-administered form (Fenson *et al.*, 1994). When MB-CDI was published in 1994, it provided a novel way of studying children from 0.8 to 2.6 of age in the 1990's (Fletcher, 2014). For the first time, it made it possible to collect large, population-representative samples from children acquiring their native language. Today, MB-CDI is possibly the most studied and widely used method for studying language acquisition within clinical settings and solely for research purposes (Behrens, 2008; Kalashnikova, 2015).

MB-CDI has proven to be a valid tool for cross-linguistic studies, and it has currently been validated for 61 languages (Dale & Penfield, 2011; Jahn-Samilo, Goodman, Bates & Sweet, 2000) and applied in large cross-linguistic studies (e.g. Luniewska *et al.*, 2015). However, the MB-CDI is not as actively studied, as different versions are being validated for different linguistic and cultural areas and used with different clinical populations (e.g. with late talkers, Heilmann, Ellis Weismer, Evan & Hollar, 2005; children with autism, Luyster, Qiu, Lopez & Lord, 2007).

Previously, MB-CDI parent reports have yielded high concurrent validity with direct language assessments (Sachse & Von Suchodoletz, 2008). In Finland, as CDEV, the MB-CDI was also validated with the Jyväskylä Longitudinal Study of Dyslexia (Lyytinen, 1999), and the use of the Finnish version of MB-CDI has been active both in clinical practices and in academic research. In the validation study, the Finnish version of the MB-CDI has been reported to correlate with RDLS production and comprehension, Bayley Scales of Infant Development (BSID), BSID mental development index (BSID MDI), with the Boston Naming Test (BNT), and with a non-word test on inflectional morphology in Finnish *Morfologia-testi* (see Lyytinen, 2003) for children at risk for dyslexia (Lyytinen, 1999). In addition to the normative study, we have gained information with MB-CDI, for example, about the early lexical and grammatical development of VLBW children acquiring Finnish (e.g. Stolt, Haataja, Lapinleimu & Lehtonen, 2008; 2009a&b), about the associations between gestures and later language with VLBW and FT controls (Stolt, Lind, Matomäki, Haataja, Lapinleimu & Lehtonen, 2016), the relations between vocabulary at 2.0 and picture-naming capabilities at 5.0 (Stolt *et al.*, 2012), and about the growth and predictive value of MB-CDI-measured vocabulary and morphological skills in children with and without familial risk for dyslexia (Lyytinen & Lyytinen, 2004).

The internal validity of MB-CDI is reported to be satisfactory and earlier measures have been presented to predict later measures in US-English, accounting for 16.6 to 31.1 percent of variance ($p<.0001$) (Bates, Dale & Thal, 1995: 101). In addition, in the normative study for Finnish, there were statistically significant associations within vocabulary scores, morphology, and MSL scores at different ages (Lyytinen, 1999). However, no internal correlational analyses were reported between the MB-CDI vocabulary and language measures.

Although there is an extensive amount of research to justify MB-CDI validity and reliability, it has also faced critique. For example, Feldman, Dollaghan, Campbell, Kurs-Lasky, Janosky and Paradise (2000) have critiqued MB-CDI for having too much variance, for not being stable, nor sensitive enough to predict later language outcomes, while Fenson, Bates, Dale, Goodman, Reznick and Thal, (2000) consider the variation to present the true nature of large individual differences. In addition, some questions about MB-CDI's suitability for comparing groups from different socioeconomic backgrounds, profiling individual children and evaluating the effectiveness of interventions have been posed (Ambridge & Rowland, 2013, review; Feldman *et al.*, 2000); MB-CDI language scores have been reported to vary as a function of sociodemographic variable in the US study by Feldman and colleagues (2000), but this finding was not replicated in the study conducted on concurrent validity of MB-CDI German version (Sachse & Von Suchodoletz, 2007). In addition, the study of Feldman *et al.* (2000) found that girls scored higher than boys, and that ethnicity was significantly associated with MB-CDI score. However, the authors pointed out that gender and race accounted for only a small amount of variance.

### 2.3.3 Digitalization and technical advancements in analyzing naturalistic data

In celebration of the 40th year of the Journal of Child Language, Dan Slobin (2014) revised the development of the field of child language studies from the 1960's to the present time and stated that three notable tools that have changed our understanding of child language development are the audio-recorder, the video-recorder, and the computer. When the recording devices were put to use, they made it possible to conduct analyses on transcribed data and to observe the reliability of observer interpretations (Behrens, 2008). This was a substantial advancement to the hand-written diary notes.

In the modern era of digitalization and advanced experimental methods, diary studies may at first glance seem to be ancient history. However, as was discussed in subchapter 2.3.1., the pen and paper approach very much continues in clinical practices in the hands of speech and language therapists, who are dealing with questions related to the delayed or disturbed phenomena of child language acquisition. In the following pages, I will summarize the technical advancements benefitting child language researchers, who are conducting research on naturalistic data (and thus, exclude experimental methods, such as nearinfrared spectroscopy, eye-tracking etc.), but also, a few technical solutions, which might be useful for clinical practices, if validated for Finnish.

The digitalization of natural samples and language acquisition studies largely emerged in the 20th century with the groundbreaking work of Brian MacWhinney (2000) – a founder and a driving force of Child Language Data Exchange system, the CHILDES-database. CHILDES was the first open-source database in the field and created the venue for open and co-operative research culture. A substantial amount of work has already been conducted with the CHILDES corpora and the automated tools within, which continue to be in active use in the current research (e.g. Che, Alacorn, Yannaco & Brooks, 2016; Goodman *et al.*, 2008; Hills, 2013). The benefits of CHILDES include cost-effective use of data because although data collection and especially transcription can still be very time consuming, the resulting dataset can be used multiple times both by the original researchers and by others, if the data is made publicly available (Ambridge & Rowland, 2013, review).

Besides CHILDES, there are also other language sample analysis (LSA) tools, which are used by clinicians and researchers abroad. These include the Systematic analysis of Language Transcripts (SALT; http://saltsoftware.com/blog/page/1/) and Computerized Profiling (CP; http://www.computerizedprofiling.org/index.html). CHILDES, SALT, and CP all make it possible to calculate many, if not all, ratios describing language proficiency listed in 2.3.1. However, both SALT and CP have been developed in the US and are mainly used within English dialects, although SALT has also been adapted for Turkish (Acarlar & Johnston, 2009) and in part for French, and CP has been used, for example, for Chinese, according to the developer's website. To the author's best knowledge, the insight here is that the CLAN annotation tool developed for CHILDES has been used in Finland and with Finnish data, but there are no Finnish corpora in the CHILDES database. However, SALT and CP have not been tested with Finnish.

In addition to already presented assets, there are also other resources that make the effort to collaborate and use data as ecologically as possible (e.g. the CLARIN consortium, Hinrichs & Krauwer, 2014). In fact, as stated by Behrens (2008), we have seen a massive proliferation of publicly available corpora in the recent years, and new ones are emerging every so often. This is important because some phenomena in child language research are relatively infrequent, and thus, are hard to study without the availability of the archived transcript data, automated search, and analysis procedures (Snow, 1995). For example, the researchers working with computational methods and natural extensive and intensive data have conducted cooperative work to establish new repositories (e.g. the newly established HOMEBANK, VanDam, Warlaumont, Bergelson, Cristia, Soderstrom, DePalma & MacWhinney, 2016), to share codes (e.g. through Github), and to offer shared venue and resources for common work (e.g. the website and meetings of the DARCLE-group, initiated by Alex Cristia). In addition, the formerly known CLEX-database, consisting of data gathered with MB-CDI, has been renewed. The new version of CLEX, nowadays known as WORDBANK (Frank, Braginsky, Yurovsky & Marchman, in revision), is an open database for researchers and offers possibilities, for example, for cross-linguistic studies.

The current technical advancements in computer technology and digital data gathering tools have only recently exponentially increased the possibilities of large data collection, storage, and handling. The abovementioned advancements have provided the opportunity to see the shortcomings of past research. For example, although the CHILDES database has been a groundbreaking innovation in child language research, the data within CHILDES were gathered largely in the era of tape recorders and limited possibilities of recording digital samples; thus, the analyses and conclusions must take sampling constraints into account. According to Ambridge and Rowland (2013, review), typical recording regimes have not been more frequent than about 1h/week with spontaneous speech samples; thus, the research has not been able to provide a direct picture of what children have heard and said. Additionally, Adolph, Robinson, Young, and Gill-Alvarez (2008) criticized past research as having had decades of reliance on cross-sectional designs and broad-sweeping longitudinal approaches, which "have left researchers with a gallery of before and after snapshots, studio portraits of newborns, and fossilized milestones, but little understanding of the process of development itself". Adolph and colleagues (2008) go on further to suggest that "what we need are accurate, fine-grained depictions of developmental trajectories for cognitive, language, perceptual, motor, and social skills".

However, in child language development, collecting large, dense, and representative samples is not easy. So far, the most extensive data have been gathered in a case-study by the Human Speechome Project (Roy *et al.*, 2006). The Human Speechome was aimed at recording everything from family home during the first years of a child's life and to develop a computational model of a child's experiences that affect language learning. To succeed, researchers of the project developed several technical solutions for data gathering, mining, and modeling and was, in fact, successful in developing a CDS detector for the project's needs and incorporating information from spatial, temporal, and extralinguistic features into their models and showing their meaning for word acquisition (Roy *et al.*, 2015; Vosoughi & Roy, 2012). However, much of the state-of-the-art tools in Human Speechome were not developed for clinical practices and are limitedly being implemented in other areas of research more broadly (however, see the wearable data gathering tool; Chin, Vosoughi, Goodwin, Roy & Naigles, 2013). Therefore, easier solutions to acquire and analyze representative data are needed.

#### 2.3.3.1 Description of the LENA™ System

The LENA Pro system is designed for researchers, speech language pathologists, audiologists, and pediatricians (LENA Foundation, 2014). It includes computer software and a recording device, a digital language processor (later DLP), which can record up to 16 hours of acoustic data in 16100 KHz sampling rate in MP3 format (Ford, Baer, Xu, Yapanel & Gray, 2008). Recordings can be organized and analyzed using the LENA System software (later LENA), a speech recognition program that uses prosodic information and statistical probabilities to segment, identify, and label speech and other audio sounds. In addition to audio analysis, LENA includes a Developmental Snapshot (see further Gilkerson & Richards, 2008; LTR-07-02), a questionnaire screening for children's developmental skills. However, as the Developmental Snapshot is not used in this study, further chapters focus on describing the characteristics of the audio analysis and reviewing past research, which has evaluated or implemented LENA analyses into child and family data.

Labelling in LENA relies on speaker segment identification and depends on algorithms, which have been derived from a large human-transcribed training set from American-English (AE) data. The software generates segmented and speaker-identified data (key child, other child, female adult, male adult, non-speech sounds, e.g. noise, electronic devices), from which it counts the amount of child vocalizations (CVC), adult words (AWC), conversational turns (CTC), automatic vocalization assessment (AVA), and different environmental sounds (see LENA Foundation, 2008). In addition, LENA software has an external data-mining tool (ADEX) to help manage large data sets and provide more detailed analyses such as the separation of female and male word counts.

Since the LENA System provides a fresh approach to child language research, several studies have used it within the US in different clinical groups such as hearing-impaired and autistic children (Aragon & Yoshinaga-Itano, 2012; Warren *et al.*, 2010). In addition, some pilot studies have explored its use for other purposes, for example, for studies on classroom interaction (Wang, Pan, Miller & Cortina, 2014), the auditory environment of elderly people (Li, Vikani, Harris & Lin, 2014), and children cared at neo-natal intensive units (Caskey, Stephens, Tucker & Vohr, 2011).

LENA Foundation and Infoture Inc. have published reports about the standardization process, validity, and reliability of LENA in several technical reports (e.g. Xu, Yapanel & Gray, 2009; Ford, Baer, Xu, Yapanel & Gray, 2008; Gilkerson, Coulter & Richards, 2008; Xu, Yapanel, Gray & Baer, 2008). All technical reports are available for anyone interested at the LENA Foundation website (http://www.lenafoundation.org/customer-resources/technical-reports/). In addition to LENA technical reports, there are also peer-reviewed publications, which have included different settings to evaluate the reliability and validity of LENA, mainly using English data.

However, the LENA System is currently being validated only for American-English, although some of its features have also been studied with the Chinese Mandarin dialect (Gilkerson, Zhang, Xu, Richards, Xu, Jiang, Harnsberger & Topping, 2014), American Spanish (Weisleder & Fernald, 2013), and European French (Canault, Le Normand, Foudil, Loundon & Thai-Van, 2015). There is, however, a growing interest in using LENA in child language studies both worldwide (see for example Pae, 2013, for Korean; Zhang *et al.*, 2015 for Chinese SDM pilot intervention) and within North American non-English and multilingual language communities (e.g. Weisleder & Fernald, 2013). Many research designs have already been displayed in conferences for environments other than English (e.g. Aldosari, Almuslamani, Wilson & Gilkerson, 2012; Ganek & Eriks-Brophy, 2013; Jackson, Callender, Diehm & Meissier, 2013; Löfkvist, 2016). Table 2 presents key findings from published peer-reviewed papers that have addressed the reliability and validity of LENA. These findings will be discussed in detail in chapter 2.3.3.2.

**Table 2.** Studies reporting LENA validity and/or reliability

| Reference | Focus of the study | Subjects | Language(s) | How the validity and reliability of LENA measures was evaluated | Results on the validity and reliability of LENA measures |
|---|---|---|---|---|---|
| Canault *et al.*, 2015 | To study the reliability of LENA measures in European French | 18 children from 3 to 48 months of age | French (native) | Reliability: From a 54-hour sample, CVC and AWC count correlations between human and LENA were reported as a whole and in age groups of 0-6, 7-12, 13-18, 19-24, 25-36, and 37-48 months. | AWC $r_s$=.64, p<.001 (correlations varied between age groups: $r_s$= .61 - .87, p<.001) CVC $r_s$=.71, p<.001 (correlations varied between age groups: $r_s$=.39 - .83, p<.001) |
| Dykstra *et al.*, 2012 | To evaluate the use of LENA in preschool classrooms of children with ASD | 40 children with ASD | Not specified (English environment) | Validity: CVC, CTC, and AWC were compared with PLS-4, The Mullen Scales of Early Learning, and ADOS. | Significant correlation between AWC – Mullen VR r=.41 and PLS-4 r=.35. CV – PLS-4 r=.33, CT – Mullen-VR r=.33, ADOS severity did not correlate with LENA measures. |
| Gilkerson *et al.*, 2014 | To evaluate LENA performance from Chinese data | 22 children and families from the Shanghai area | Chinese SDM (Shanghai Dialect and Mandarin) | Three 5-min samples (2% of total recordings = 5.5 hours) with high CT were transcribed and compared with LENA estimates. Segmentation accuracy, sensitivity, and precision were studied for AWC, CVC, and CT counts and discrimination agreement in child speech and non-speech sounds was compared. | Sensitivity was good for adult (79%) and child (81%) segments, precision for adults 66%, but 27% for children. Discrimination of speech-sounds from non-speech sounds was good for sensitivity (84%) and precision (77%). AWC did not differ substantially between LENA and human estimates. |
| Greenwood *et al.*, 2011 | To study LENA measures using middle to upper SES TD families | 30 TD children and their families | English | Validity was assessed via cross-sectional, longitudinal research design and included traditional measures (BSID-III, PLS-4, STAR) and LENA analysis. | LENA and BSID-III, AWC, and PLS-4 did not correlate. PLS-4 correlated with CVC (PLS-4total, r=.51, p<.01) and CT (PLS-4total r=.50, p<.01). |

| Reference | Focus of the study | Subjects | Language(s) | How the validity and reliability of LENA measures was evaluated | Results on the validity and reliability of LENA measures |
|---|---|---|---|---|---|
| Oetting *et al.*, 2009 | To evaluate the device and analysis and explore their potential | 17 child-mother dyads (from tape) | English | Inter-rater-agreement was conducted by comparing results of the analysis with the results from transcriptions done using SALT and the stability evaluation of the LENA System by doing two passes of the same recording. | The 1st (r=.85, p<0.001) and 2nd (r=.71, p <0.001) pass correlated significantly. AWC correlated with the SALT transcription. For CVC, the correlation between the 1st and 2nd pass of LENA was r=.76, p<0.001.<br><br>For CT, transcription and the 1st and 2nd LENA passes were low (1st r.14, p<.05 2nd r=.08, p>.05), but authors pointed out differences with SALT and the LENA System, which are thought to influence the results. |
| Soderstrom & Wittebolle, 2013 | To study variation in language input in typical daily activities during a day at home and in childcare facilities | 11 TD children, ages 12-29 months | Not specified | An inter-rater reliability test was conducted as follows: 183 5-minute samples from daycare (100) and home settings (83) (computer selected) were hand-coded (two passes) and compared with the AWC and CVC estimates provided by LENA software. | AWC overall correlation r=.76, p<.001 (daycare r.77, p.<.001 and home data r=.83, p<.001). CVC overall correlation r=.67, p.<001 (daycare r=.72, p<.001 and home r=.65, p<.001.<br><br>Additionally, mean absolute errors were presented for both settings and for several types of activities |
| Warren *et al.*, 2010 | To compare language production and language environments in TD children and children with ASD | 26 ASD and 65 age-matched TD children were included | ASD not specifies, TD sample are English-speakers | Validity was assessed by comparing LENA measures to language and developmental ability questionnaires (CSBS, CDI, LDS, MB-CDI, BRIEF-P) and symptom questionnaires (M-CHAT, SCQ, CBCL). | Higher AWC and CT count was associated with reduced symptom levels and increased language (and/or) ability scores. Positive correlation was highest for CT and MB-CDI verbal production (r=.80, p<.01), CT and CDI total (r=.78, p<.01), and CT and CSBS total (r=.76, p<.01). The severity of symptoms was negatively correlated with M-CHAT and AWC (r=-.66, p<.01), CT and SCQ (r=-.57, p<.01) and M-CHAT with CT (r=-.52, p<.01) |

| Reference | Focus of the study | Subjects | Language(s) | How the validity and reliability of LENA measures was evaluated | Results on the validity and reliability of LENA measures |
|---|---|---|---|---|---|
| Weisleder & Fernald, 2013 | How the amount of speech affected children's language processing and vocabularies | 23 infants | Spanish | An inter-rater reliability test for AWC was conducted from ten 60-minute samples by a native Spanish-speaking coder. | Inter-rater agreement for AWC was high (r=.80). |
| Zimmermann *et al.*, 2009 | Testing the independent association of adult language input, television viewing, and adult-child conversations with language acquisition | phase 1: 275 families, phase 2: 71 families | English | Inter-rater agreement was studied from 70 human-coded, 12-hour sessions that were compared with those reported from the software. Validity for AWC and CT count was assessed by comparing counts to PLS-4. | A substantial to high agreement (70.5 – 82%) rates were reported for AWC, CVC, TV, and other sounds. Cohen's kappa for adult speech between a human coder and LENA was $\kappa$=0.65 and for TV $\kappa$= 0.57. Higher AWC and CT levels were reported to associate with increased PLS-4 scores and television exposure – to decrease in the language score. |

#### 2.3.3.2 Reliability and validity of LENA speaker identification and core counts

The LENA System uses pre-defined rules for segmenting audio stream and American-English-based (AE) probabilistic models to identify and label sound segments with speaker labels (key child near/far, female adult near/far, male adult near/far, other child near/far) or labels for environmental sounds (overlapping near/far, noise near/far, electronics near/far, and silence). The role of segmentation accuracy and correct labelling is crucial for LENA adult word count, child vocalization count, and conversational turn count measures, as these counts are grounded in segmentation, speaker identification and phone recognition.

Inter-rater percent agreement (between LENA and "human", "human" considered the gold standard) for speaker identification from the normative sample has been presented in the LENA technical report (LTR-05-02) and is reported to be 82% for adult, 76% for child, 71% for TV, and 76% for other sounds (Xu, Yapanel & Gray, 2008). However, percent agreement has been widely criticized, as it includes only the observed agreement, but fails to take chance into account (Hayes & Hatch, 1999). Therefore, for the purposes of examining observer agreement, it would be advisable to use, for example, Cohen's kappa (ĸ) (Viera & Garrett, 2005). In addition, for any potential diagnostic tools, the diagnostic accuracy should also be tested, for example, using sensitivity, specificity, overall accuracy, and predictive and/or discriminative values (Eusebi, 2013; Okeh & Okoro, 2012, review). VanDam and Silbert (2013b) have further stated that an important goal of automatic labelling is to maintain relatively high precision by reducing false positives. Studies conducted on LENA reliability that have looked beyond agreement rates are summarized below.

In the studies of VanDam and Silbert (2013a) and Oller and colleagues (2010), percent agreement between machine-coded segment labelling and human judges was counted, but these were reported in addition to kappa-statistics. VanDam and Silbert (2013a) found percent agreement to be higher for children (85.9%, Cohen's kappa ĸ=.708), but lower for male adults (60.9%, ĸ=.599) and female adults (59.4%, ĸ=.503), when compared with LENA Foundation's agreement rates. Oller and collegues (2010) followed previous studies and chose to study agreement rates for child versus adult segments, which were found to be 73% with 5% of false positives (when "human" was used as the gold standard). However, Gilkerson *et al.* (2014) chose to explore the ability of LENA to identify speakers from Chinese Mandarin dialect data through sensitivity (true positives from true positives and false negatives) and precision (true positives from true and false positives, also called "positive predictive power"). Gilkerson and colleagues found that LENA showed to be similarly sensitive to child and adult segments as in AE validation, but precision in child segment identification was found to be poor.

The reliability of the LENA System counts has been considered a part of studies conducted with typically developing children (TD), late speakers (LT), and children with autism spectrum disorder (ASD), and mainly with English-speaking populations for adult word count (AWC) and child vocalization count (CVC) (Table 2). All reliability tests have been conducted by comparing LENA segments, counts, and estimates with ones provided by human transcribers. For AWC, inter-rater correlations between LENA and human coders have been reported to correlate between *r*=.76 and *r*=.83, respectively, and, more importantly, encouraging results have been reported by Spanish (AWC *r*=.80) and Chinese SDM studies (AWC to SDM orthographic words *r*=.73, p<.001) (Gilkerson *et al.*, 2014; Weisleder & Fernald, 2013). For CVC, inter-rater agreement has been reported to range from *r*=.65 to *r*=.76. However, to the author's best knowledge, it seems that all LENA core measure reliability tests have so far been conducted with correlative analyses, which may not be the most reliable way in conducting such research (Bland & Altman, 1986; Haber & Barnhardt, 2006). Bland and Altman (1986) have stated that the use of the correlation coefficient is inappropriate in agreement studies, for example, because a high correlation coefficient does not actually mean that the two measurements agree, but also that data that seems to be in poor agreement can produce high correlations. In addition, the author is not aware of any studies that would have questioned LENA's ability to distinguish multiple child speakers from each other.

LENA AWC, CVC, and CTC (conversational turn count) measures have also been compared to various types of language, social behavior, and developmental measures. AWC has been reported to correlate positively with the Preschool Language Scale (PLS-4) scores (*r*=.35, *p*<.05), Mullen Scales of early Learning (MULLEN-VR; *r*=.41, *p*<.01) (Dykstra, Sabatos-DeVito, Irvin, Boyd, Hume & Odom, 2012) and negatively with increased scores from The Modified Checklist for Autism in Toddlers (M-CHAT) (*r*=-.66, *p*<.01) (Warren *et al.*, 2010). CVC has been reported to correlate positively with PLS-4 (*r*=.33-.51, *p*<.01) (Greenwood, Thiemann-Bourque, Walker, Buzhardt & Gilkerson, 2011).

LENA CTC has been studied in relation with information about children's performance in traditional measures and/or parent reports. LENA CTC has been reported to correlate positively with PLS-4 (Greenwood *et al.*, 2011, see also Dykstra, 2012, for close to significant correlation), MULLEN-VR (*r*=.33, *p*<.05) (Dykstra *et al.*, 2012), Communication and symbolic behavior scales (CBCS) (*r*=.76, *p*<.01), The Child Development Inventory (CDI) (*r*=.78, *p*<.01), and The MacArthur-Bates Communicative Inventory (MB-CDI) (*r*=.80, *p*<.01) (Warren *et al.*, 2010). Negative correlations have been found between CTC and several tools screening for atypical behaviors. CTC correlated statistically significantly with M-CHAT ($r$=-.52, $p$<.01), The Child Behavior Checklist (CBCL) ($r$=-.39, $p$<.01), and the Social Communication Questionnaire (SCQ) ($r$=.-57, $p$<.05) (Warren *et al.*, 2010). However, the Autism Diagnostic Observation Schedule (ADOS) did not correlate with LENA measures in the study of Dykstra *et al.*, (2012), but the authors suggest that the result may reflect the small sample size of the study. An older version of LENA (V 2.3.) has also correlated positively with SALT transcription for AWC ($r$= .71-.85, $p$<.001) and CVC ($r$=.76), but not for CTC (Oetting *et al.*, 2009).

# 3 Aims of the study

It is not known how the early language of twins develops, and what the role of biomedical and social environmental variables is in their language development. In addition, it is only because of recent technological advancements that it has become possible to study naturalistic social interaction without sampling restrictions as it is occurring in families living their daily lives and to discover the very basic information needed to understand language acquisition through socialization. Therefore, this study is two-fold in nature, relating to a) questions about the reliability of automated technology and its performance in relation to traditional parental questionnaires, and b) questions about twins' language development and the role of the social, pre-, and neonatal environment in language development. Firstly, this study aims to explore whether the algorithm of the automated method provides reliable information about the detection and identification of speakers and the accuracy of child utterance and adult word counts. Secondly, the automated method is applied to measure the quantity of speech and speech-like utterances spoken in twin families and to explore, if children neonatal health and demographic variables have any effect on the volubility of different family members. Thirdly the study aims to discover how babbling and early linguistic skills develop in twins and to explore if the neonatal health and demographic variables affect development. Lastly, the study aims to discover whether there are associations between variables of quantified family speech and parent reported variables of twins' language development.

In the first part of this study, the reliability of a novel method and its automated analysis (Language Environment Analysis™, later LENA) is assessed with a special focus on segmentation accuracy, speaker identification, and reliability of adult word and child vocalization counts. These are studied using the following questions:

1. How similarly does LENA and a native Finnish-speaker identify speakers?
2. How reliably does LENA identify key child vocalizations in key child segments from non-vocal elements (i.e. cries and vegetative sounds) compared with human-identified vocalizations?
3. How accurate are LENA-provided adult word counts (AWC) and child vocalization counts (CVC) compared with counts provided by native Finnish-speakers?

Secondly, automatic LENA analyses are utilized to gain an understanding of what twin children hear during their typical day. In addition, the second part also inspects the amount of vocalization produced by twins and the possible relations of shared and non-shared environmental variables to the quantity of vocalization and input frequency in twin families. These themes are addressed using the following questions:

4. How much families talk, according to LENA spoken segment durations of key child, other child, male, and female adult?
5. How much does LENA suggest children to hear adult words, participate in conversational turns, and produce child vocalizations?
6. Do differences in social and biomedical environments affect the amount of automatically detected speech and vocalizations?

The third part inspects twin children's language development through the eyes of the parents: how parents discover vocal milestones from pre-lexical stages, and how children acquire vocabulary and language skills in early toddlerhood. These themes are addressed through the following questions:

7. When do parents report their children starting reduplicative and variegated babbling and when do they discover their children's first protowords?
8. How do twin children's vocabularies and language skills develop during the second year of their lives, when compared with normative information?
9. Does the emergence of vocal milestones, the acquisition of vocabulary, and language skills differ when social and biomedical environmental differences are compared?

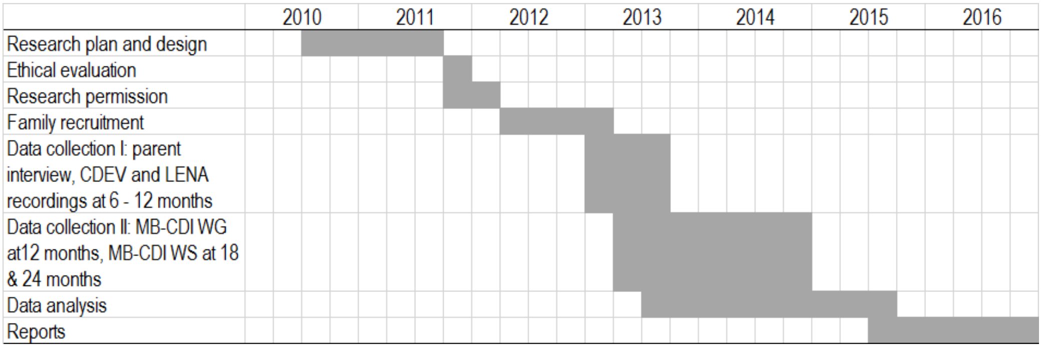
In the final part, the associations of LENA-measured heard input and the parent-reported onset of vocal milestones, the quantity of children's vocabulary and language development are studied. These themes are addressed using the following questions:

10. Is the amount of LENA-detected speech or speech-like vocalizations associated with the LENA-detected volubility of family members?
11. Is there a relationship between the LENA-detected amount of child vocalizations and heard input with the information gathered from pre-lexical development, vocabulary development and language development using parent questionnaires?
12. Is there a relationship between parent-reported vocal milestones, early vocabulary, and language skills in twins?

# 4 Subjects and methodology

## 4.1 Study design and research procedure

This study is a longitudinal study with monthly recordings and follow-ups of vocal milestones of twins at 6 – 12 months and a follow-up of vocabulary and language development up to two years of age (Figure 1). The study is exploratory by nature, in the sense of applying new automated methods to Finnish data. For the analyses on language environment and language development, the work used descriptive statistics, applied non-parametric hypothesis testing, and non-parametric correlative statistical methods.




CDEV = Checklist for vocal development
MB-CDI WG = MacArthur-Bates Communicative Inventories, Words and gestures
MB-CDI WS = MacArthur-Bates Communicative Inventories, Words and Sentences

**Figure 1.** Flow chart of the study process

## 4.2 Ethical commitments

Logopedic research combines humanities, behavioral, and medical research traditions and is obliged to conduct research with high ethical standards, following ethical procedures applied for both human and medical sciences. To ensure high ethical quality, the research plan of this study was sent to the chairperson of the Regional Ethics Committee in Pirkanmaa Hospital district for a preliminary review. This procedure confirmed that no further evaluation was needed. However, before the recruitment process via public child health clinics, additional permits were requested from the collaborating cities. All permissions are placed in the archives of logopedics in the University of Tampere and are available for inspection.

Many of the logopedic phenomena are often at least partly language-specific, and informants are recruited from different sub-groups, which are formed, for example, on the basis of a medical condition or a developmental trait. Therefore, the groups suitable to act as informants are often few in number and the anonymity of participants may need to be carefully protected. This was also the case in this study, as twin families are few in number when compared to the Finnish population; furthermore, they often also know each other, for example, due to birth and parenting coaching and activities arranged by the Finnish Multiple Birth Association. In this study, the aspect of confidentiality was taken into account when considering the background information to be reported.

Volunteer families participating in this study lived in the Western part of Finland and within several hospital districts. Before the onset of the study, the author discussed the research protocol with the families, and written information handouts (see Appendix 22) were also given out. In addition, time was reserved for questions arising from the members of the participating families. Consent forms (Appendix 23) were filled out and families had the possibility to restrict the further use of research material, if they wanted. If families did not decide otherwise, the expiration date for data storage was set to the date when the children turn 18. In addition, families were informed that there was no fee paid to participants, but all families were promised to receive developmental profiles of their children's early language development. Families were also informed of the possibility to withdraw from the study at any time and to ask information about their children's profiles during the whole research time, if needed, for example, for evaluation of developmental and/or medical problems, or for the design of rehabilitation services.

As this study was conducted with families going through a potentially demanding period with twin babies, special emphasis was paid to designing data collection as feasible as possible for the participating children and families. This meant that the researcher delivered the data collection devices (see 4.3) first hand to the families and collected the devices from family homes after the recorded day.

In Finland, logopedic studies often face the challenge of balancing between sample sizes, representativeness, and conducting data collection within a reasonable time period. In this sense, this study is no exception. The small group of children (N=20) and their families (N=10) cannot be seen to represent Finnish twin population in the sense that we could make extensions beyond the studied group. As in any recordings, the subjects in this study were of course obliged to choose what they did and spoke during the recordings, and if needed, they were also allowed to pause the recording device for a short period of time.

## 4.3 Subjects and data collection

After receiving study permissions, the recruitment of twin families began. Families were recruited via research website (http://kaksostenpuhe.weebly.com/kaksosten-jokeltelun-ja-varhaisen-sanaston-kehitys-tutkimus.html) and other websites directed at parents (e.g. Vauva.fi, Monikko-onnea discussion forum), leaflets delivered to child health clinics, and with the help of the Finnish Multiple Births Association (Appendix 24). All of the volunteering families contacted the researcher first, and at that time, a preliminary interview about severe medical and developmental conditions was conducted by the author, as they were set as exclusion criteria to participation. One pair of twins was excluded due to a diagnosed syndrome affecting speech and language development. After exclusions, twenty twin participants (F 12, M 8) and their families (N=10) living in the Western part of Finland were recruited and formed the final sample.

When clinicians meet twins with a need for language evaluation, they need to systematically explore all possible reasons that might account for children's speech and language development to be able to find appropriate case-based solutions (Mogford-Bevan, 2000). In this study, the information about the family background and pre- and perinatal factors was obtained through semi-structured parent interviews following the questionnaire form in Appendix 21.

From the ten participating twin-pairs, one pair was reported as a monozygotic pair, one pair as unknown, and 8 pairs as dizygotic twins. Thus, the comparison between mono- and dizygotic twins was not feasible (see Table 3 for a list of background variables). Of the twenty children, 12 children were born as full-term after 37 weeks of gestational age (GA), whereas eight children were born as late preterms (34 – 36 of GA). Eight of the children were born low-birth-weight (LBW), but as only one child was found to be born small for gestational age (SGA), no comparison with SGA and children with appropriate for gestational age (AGA) could be conducted.

The Apgar-score of the children greatly differed for one-minute scores (2–10), but variation was more modest for five-minute scores (7–10). However, the use of Apgar-scores was not feasible because of multiple missing values, and thus, the decision was made to use the level of neonatal care as a background variable for neonatal health. In this study, all 13 children cared at bedside (rooming-in) were also the same children born by vaginal birth, and 7 children cared at the neonatal ward were born by cesarean section. In further chapters, however, to help the reader and avoid multiple analyses, these children are referred to only for the level of neonatal care.

**Table 3.** Characteristics of the participating children

| Variable | | Count |
|---|---|---|
| gender | male | 8 |
| | female | 12 |
| older siblings | no siblings | 10 |
| | older siblings | 10 |
| term | term | 12 |
| | preterm | 8 |
| birthweight | LBW | 7 |
| | NBW | 13 |
| complications at birth | no | 13 |
| | yes | 7 |
| level of neonatal care | bedside | 13 |
| | neonatal care | 7 |

Complications at birth included prolonged births, one acute asphyxia with rapid recovery, and umbilical cord and breech deliveries. However, it should be noted that the children born through complicated births are not the same children who needed neonatal care. Out of the children who needed neonatal care, two children needed antibiotic treatment after birth, and this treatment was delivered in neonatal intensive care during the first postnatal days of the twins' lives. In addition, five of the children were treated at the neonatal ward. However, for group comparisons, all seven children, who needed treatment at the ward, were treated as one group, and all 13 children cared at beside formed another group.

The parents of twins did not report their children to have any severe health problems or significant developmental problems during the period of data collection, and all the children were nurtured at home for the first year of their lives. In half of the families, twins were first-born children, whereas the other half of the participants had older siblings. From the twenty participants, fourteen children had a monolingual family background. However, in all of the three multilingual families, the mother (the primary caregiver at home) was a native speaker of Finnish.

When the educational attainment of the parents of twins in this study was compared to information from the Official Statistics of Finland (http://www.stat.fi/til/vkour/index_en.html), the parents of twins were found to be more educated than the Finnish population on average. The comparison was conducted as follows: 1) the educational attainment of 20-44-year-old females and males were derived from the official statistics, 2) the level of educational attainment was transformed into percentages (resulting in a mean of 13% of higher education in males and 17% in females), 3) the attainment of education reported by participating adults (50% of master's degree or higher in females and males) was compared to information from official statistics (see the education attainment level of the participants in the present study in Table 4).

All but one mother and one father had a degree in higher education, but parent education was dichotomized for the feasibility of group comparisons. Thus, parent education as a background variable consisted of a parent having a vocational college or a bachelor's degree and parents having master's degrees or higher.

**Table 4.** Educational attainment of the parent in twin families¨

| Degree | M | F | All |
|---|---|---|---|
| Vocational college or bachelor's degree | 5 | 5 | 10 |
| Master's degree or higher | 5 | 5 | 10 |

The data in this study consists of recordings of the twins from 6-12 months of age and from CDEV and MB-CDI questionnaires. The recordings were gathered from simultaneous monthly recordings in the homes of the participating families, when the children were in the corrected ages[4] of 6-12 months. The recordings (N=142) were collected using LENA System's digital language processors (DLP), and the recording time was set to a minimum of ten hours per recording. Thus, the total duration of all raw audio data was approximately 1500 hours. The recordings were gathered simultaneously from both twins. However, during data collection, one DLP was broken. When the broken DLP was in repair, recordings were conducted so that children from the same family wore the remaining intact device for two consecutive days. As this resulted in recordings of different lengths, all analyses were conducted with 12-hour adjusted counts of speaker segment durations and LENA core counts.

Both of the parent-filled questionnaires were selected based on the following requirements: the questionnaires needed to be standardized and normed for the Finnish population and, to make the research more applicable, the questionnaires were selected from the ones used in clinical settings in Finland. Following this criteria, the Checklist of vocal, motor, and fine motor development (later CDEV) (Lyytinen *et al.*, 2000) was selected to acquire information about the milestones of prelexical development. The Finnish version of MacArthur-Bates Communicative Development Inventories (later, MB-CDI) was selected to acquire information about the early lexicon and language development of the children.

CDEV and the instructions of its use were given to participating families during the first recording day, and its use was controlled during the monthly visits of delivering recording devices to the families. The Finnish version of MB-CDI Words and Gestures (later MB-CDI WG) was delivered with instructions a few weeks before the child's

[4] Corrected age is calculated from the full gestational age, 40 pregnancy weeks

corrected age of 12 months, and the Finnish version of MB-CDI Words and Sentences (later MB-CDI WS) and instructions were also delivered two weeks before the corrected ages 18 and 24 months. Parents received the questionnaires either during monthly visits on recording days or via mail. In the instructions, the importance of filling out the forms at the exact age of 12, 18, and 24 months of corrected age was emphasized. Families also received a return envelope to ensure that there were no expenses for participating families. The researcher was also tasked with following up on the receiving of the questionnaires and reminded families about the questionnaires after two weeks, if they had not been returned.

## 4.4 Data analyses

### 4.4.1 Assessing the reliability of the automated method

In the first part of this study, the performance of LENA speaker recognition, identification, labelling of recognized segments, and the accuracy of LENA core measures were studied with two whole-day recordings. In order to explore possible differences that might emerge from differing environments, recordings were selected by following certain differentiation criteria: one recording was drawn from families of firstborn twin children and the other from families with twins and older siblings. In addition, the recordings were chosen from different-aged children and from a girl and a boy. With these criteria, a selection of records was made. From this selection, two recordings (R1: 11:38 hours and R2: 10:30 hours) were drawn. The key child (i.e. child wearing the DLP) in R1 was at the time a seven-month-old second-born twin girl with an older sibling, and the key child in R2 was a nine-month-old first-born twin boy with no older siblings.

One trained, native Finnish human coder segmented, labeled, and transcribed the recordings by correcting the pre-segmented LENA files, first using Transcriber (C. Barras, Geoffrois, Wu & Liebermann, 2000) for a preliminary transcription and labelling, and secondly, using Praat (Boersma & Weenik, 2014) for more detailed segment boundary correction and finalizing transcription.

The correction of LENA-provided pre-segmentation was conducted as follows:

1. During the first listening conducted in Transcriber, the human transcriber listened to the recordings and corrected LENA speaker identification by segmenting undetected speaker utterances and removing false discoveries. Simultaneously, the segments were labelled according to LENA speaker categories labelling. For human voices, the coder used the labels of key child

near/far, adult female near/far, male adult near/far, other child near/far, and additionally corrected other sounds labelled in LENA as overlapping near/far, noise near/far, TV near/far, and silence.

2. In the second listening conducted in Transcriber, the human transcriber confirmed the speaker labels, labelled the key child segment to speech-like vocalizations, vegetative sounds, cries, and silences, and transcribed the recordings orthographically in all segments, where speech could be distinguished.
3. In the third listening, the human transcriber finalized the transcriptions in Praat and adjusted the placements of segment boundaries, if utterances were estimated to be significantly longer than was discovered by LENA.
4. LENA- and the human- corrected transcriptions were exported from Praat and imported to WinMerge (available from winmerge.org), where the author manually aligned the labels and transcriptions from the recordings for comparison (see figure 2.). The primary alignment criterion was the onset time stamp from each of the segments in recordings. If the files included multiple segments with the same onset time, labels of LENA and human were used as a secondary alignment criterion (see Figure 2).
5. From WinMerge, the aligned sheets from LENA and human transcriptions were exported to an Excel sheet.
6. The author manually calculated all transcribed words on the sheet and compared the calculations from human transcriptions to the calculations provided by the LENA System.

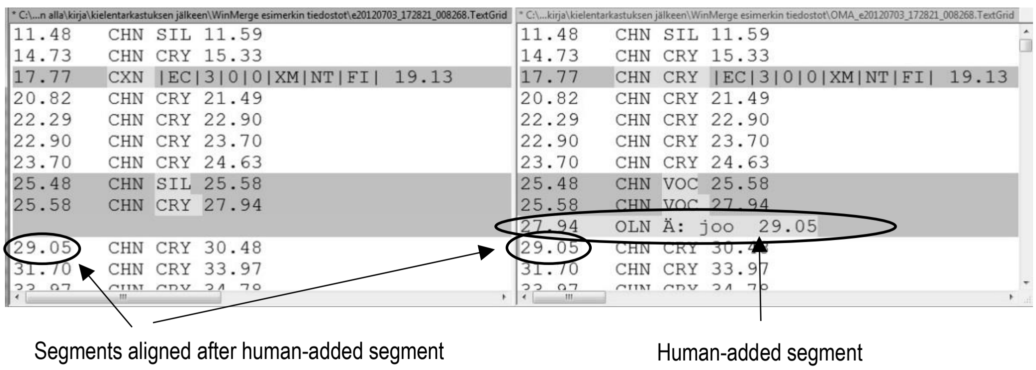



**Figure 2.** Example of segment alignment in WinMerge

Although the primary human coder used the same labels used by LENA, speaker labels were combined for analysis as follows: key child (includes key child near/far), female adult (female adult near/far), male adult (includes male adult near/far), other child (includes other child and/or children near/far). Inter-rater reliability procedures were conducted with two human coders (see below). In addition to speaker identification, how well LENA identified key child sounds as vocalization, vegetative sounds, cries, and silence was also studied.

For the label accuracy investigations of speaker identification and key child segment type identification, percent agreements and kappa-values were first calculated to ensure the possibility to compare results with previous studies. However, to gain more specific information about LENA performance, how well LENA identified the number of true positives from all positives (sensitivity), the number of true negatives from all negatives (specificity), and the number of true assessments from all assessments (accuracy) was also studied. Following VanDam and Silbert, 2013b, the importance of true predictions was emphasized, and hence, we further calculated false discovery rates (FDR) to present how many errors from LENA were type I errors, i.e. the amount of false positive predictions from the algorithm.

The experiment of the reliability test of the child and adult volubility measures CVC and AWC were executed using transcriptions from the two previously mentioned recordings, R1 and R2, from which absolute values of adult words and child vocalizations were first counted and then compared with AWC and CVC counts reported by the LENA software. Human-provided AWC was counted from human-identified female and male adult near segments because in LENA, AWC is calculated only from such segments (Xu, Yapanel, Gray & Baer, 2008). Differing from previous studies, no correlation coefficients were calculated for CVC and AWC (see 2.3.3.2. and Bland & Altman, 1986). Instead, relative error rates were calculated per hour using "human" as the gold standard. Relative error presents how large the estimation error is when compared to actual (human-provided) value and is calculated as $relative\ error = \frac{measured\ value - actual\ value}{actual\ value}$. An hour was selected as an observed time unit for relative errors, as LENA-based AWC estimate is a function of recording time, and the variability of estimation error is reported to plateau to <20% after an hour (Xu, Yapanel & Gray, 2009). The selection to study full-day recordings is also justified because the LENA technical report LTR-05-02 suggests that the error percent for AWC estimation from full-day recordings remains at approximately 5%.

In addition to human-machine inter-rater reliability, a reliability test of the assessment of the primary human coder was also conducted. For this, two trained inter-rater human coders were recruited. For the reliability test of segment labelling, the two human inter-raters labeled a 10 percent continuous sample (132 minutes) of the total duration from both of the recordings R1 and R2 with similar instructions, which were followed by the primary human coder. The labels from two additional coders were compared with the labels by the primary human coder. Human inter-rater agreement yielded a good to very good agreement. For R1, point by point agreement was 81.9% and $\kappa$ =.766, and for R2, point by point agreement was 87.4% and $\kappa$ =.791, $p<.0001$. For the human inter-rater reliability of adult word and child vocalization counts, two trained human coders transcribed 132 minutes (10% of total duration of R1 and R2) in total, and the relative error rate was calculated using inter-raters as gold standards. Relative error rates remained low for both recordings (R1 relative error rate = -.08 and R2: .28), suggesting a moderate to good reliability of the counts provided by the primary coder.

### 4.4.2 Applying automated analyses

The second part of the study focused on measuring the amount of spoken interaction near children and discovering if neonatal health and demographics would influence the amount of spoken interaction in the studied twin families. The second part of the study was hence conducted with an a priori assumption that LENA automatic analyses could be implemented in Finnish data. Although validation of the method is beyond the scope of this study, the possibility to observe LENA performance from multiple angles provides complementary information for the discussion of the suitability of the method. In addition, the results provide information on the body of results in the fields studying early language acquisition and language environment studies conducted with and without LENA.

The LENA System was used for automated audio analyses and data extraction with all 142 recordings. Table 5 presents variables measured with automated methods, which included the LENA "core measures": adult word count (AWC), child vocalization count (CVC), and conversational turns count (CTC). All LENA counts were compared with LENA norms (US), which were requested from the LENA Foundation. The reliability of AWC and CVC is reported from two case studies in chapter 5.2., but not for CTC; CTC is a measure, which builds upon speaker segmentation, adult word, and child vocalization detection (Xu, Yapanel, Gray, Gilkerson, Richards & Hansen, 2008) (see also chapters 2.3.3.2–2.3.3.4).

Besides the core measures, segment durations from LENA-identified vocalization blocks in daily recordings were extracted for "near" key child, female adult, male adult, and other child speaker segments. In LENA, "near" spoken segments are interpreted to be meaningful in the sense that near segments can be heard by the developing child. However, the "near" segments actually include only clear segments, i.e. segments with no overlapping sounds, fuzziness, or diminished sound level and are thus not absolute presentations of all spoken utterances, but offer a way to measure the proportion of clear-identified speech. Therefore, the current analyses cannot capture the amount of distant speech, as the LENA System does not consider distant speech as meaningful speech. Following Wesleder and Fernald (2013), all analyses were conducted with values of 12-hour estimations to make analyses from different length recordings compatible.

**Table 5.** Language environment variables from automated LENA analyses

| | |
|---|---|
| The amount of daily meaningful speech | Female adult segment durations adjusted to 12 hours<br>Male adult segment durations adjusted to 12 hours<br>Key child segment durations adjusted to 12 hours<br>Other child segment durations adjusted to 12 hours |
| LENA core measures | Adult word counts adjusted to 12 hours<br>Child vocalization counts adjusted to 12 hours<br>Conversational turn counts adjusted to 12 hours |

LENA-provided information about the language environment and volubility of twins was analyzed with inferential and descriptive statistics using mean (*M*) and standard deviation (*SD*), but also median (*Mdn*) and interquartile range (*IQR*) to present information about the dispersion of the data. In addition, data dispersion is presented with visualizations of repeated measures within the recording period.

As the audio data was available from seven measure points from 6–12-month recordings, the Friedman test of cross-sectional independence was used to compare the similarity of variance in different points of measurement within language environment variables. The relations of background variables (presented in 4.3) and all LENA-provided volubility and core measures were studied through group comparisons using the non-parametric Mann-Whitney U-test. In both the statistical tests, exact $p$-values were used, and the statistical significance was set to $p=.05$.

The Mann-Whitney U-tests were conducted both on monthly points of measure and on the whole pooled data. This procedure was selected because 1) monthly comparisons may reveal if any of the background variables have consistent influences on child and adult volubility, but 2) the comparison conducted with the whole data may reveal if the monthly data contained unimportant findings, as they could be washed away in the observations in the whole data. In addition, 3) the monthly comparisons may not show small but significant differences when accumulated that may have their origins in selected background variables.

To study the impact of selected background variables to child volubility and language environment of the twins, the first step was to ensure the relevance of background variables. For this purpose, a principal component analysis was conducted. As all background variables presented in 4.3. were seen to have importance for the data, no background variables were excluded from analysis. Therefore, the significance testing was conducted for child neonatal health-related background variables (birth weight, pretermity, complications at birth, and neonatal care) and for demographic background variables (birth order, gender, parent education, and older siblings) for all automated measures using the Mann-Whitney U-test. In both cases, the magnitude of difference between groups is presented with effect sizes that were calculated with the Wendt's formula of rank-biserial correlation ($r_{rb} = 1 - (\frac{2U}{n1*n2})$ (see e.g. Kerby, 2014). However, due the limited sample size, the effect sizes of background variables on dependent variables is discussed only for the group differences evident in the comparisons conducted with whole pooled data and not for monthly group comparisons. This selection enhances the reliability of interpretations, as unimportant findings may be diminished and small but important findings may accumulate on the whole data level.

### 4.4.3 Analysis of developmental milestones and trajectories

In the third part, the information about CDEV vocal developmental milestones (reduplicative and variegated babbling, protowords) and the development of vocabulary and early language skills reported with MB-CDI are reported using descriptive statistics and comparing the information about their central decencies and dispersion to normative information reported in method manuals (see Table 6 for list of variables).

**Table 6.** Variables of prelexical milestones and trajectories, as defined by CDEV and MB-CDI

| | |
|---|---|
| Vocal milestones | the onset of reduplicated babbling |
| | the onset of variegated babbling |
| | the appearance of protowords |
| Vocabulary | receptive vocabulary at 12 months |
| | expressive vocabulary at 12 months |
| | expressive vocabulary at 18 months |
| | expressive vocabulary at 24 months |
| Early cognitive-linguistic skills | first signs of understanding at 12 months |
| | the sum variable of actions and gestures at 12 months |
| | the sum variable of morphology at 18 and 24 months |
| | use of words at 18 and 24 months |
| | MSL at 18 and 24 months |

The MB-CDI words and gestures-form (WG) was used to gather data from twins at the corrected ages of 12 months, and the MB-CDI words and sentences-form (WS) was used at the ages of 18 and 24 months. As only the WG version separated receptive and productive vocabularies, only expressive vocabularies were available from 18 and 24-month-old twins. In this study, the parent reporting first signs of understanding was used as a general measure of the emergence of linguistic understanding of the child and the sum variable of actions and gestures – as a measure of non-verbal communicative abilities of the child. In addition, the MB-CDI's use of words-measure was applied as a proxy of the child's ability to use and understand referential language, while the sum variable of morphology was selected to reflect the child's ability to produce inflectional morphological forms of their first language. Additionally, the maximum sentence length (MSL) of three longest expressions was selected as a measure to represent a child's capability to formulate meaningful syntactic-grammatical expressions.

Although the current work applies non-parametric methods, the descriptive information in addition to the median and the inter-quartile range also includes information about mean and standard deviations of the selected sample. These are presented to ensure that the reader will be able to compare information from twins to the information presented in the CDEV and MB-CDI manuals. As the role of an older sibling in previous literature has been seen as an important factor influencing family interaction, a comparison with non-parametric methods is conducted between first-born twins and twins with older siblings. In addition, data is also compared for other possibly influential background variables, including demographics and factors of neonatal health (see theoretical basis in chapter 2.1. and chapter 4.3 for a description of background variables). The inclusion of selected neonatal health and demographic variables was also supported by principal component analysis conducted prior to data analysis.

All group comparisons were conducted using the non-parametric Mann-Whitney U-test, and the magnitude of difference between groups is presented with effect sizes calculated with the Wendt's formula of rank-biserial correlation (see formula in 4.4.2.) (Kerby, 2014). In group comparisons, exact *p*-values were used, and the statistical significance was set to 5% ($p$=.05). The analysis of the associations between vocal milestones and measures of vocabulary and language skills is conducted by using non-parametric correlations Spearman rank correlation coefficient ($r_s$).

### 4.4.4 Uniformity of information from automated and checklist measures

The fourth part of the study focused on studying uniformity of the information gathered by automated methods and with the two checklists selected for this study. Previously, Lyytinen and colleagues (1996) have reported that parent-reported milestones of prelinguistic development in CDEV were correlated with the Finnish version of MB-CDI scores, and the concurrent validity of MB-CDI has been widely reported (see reliability reviews in 2.3.2.2 and 2.3.2.4). In addition, LENA measures have previously been compared to traditional clinical measures for children acquiring English (see review of LENA reliability in 2.3.3.2-2.3.3.5), but to the author's best knowledge, none such research have been conducted for non-English language areas.

The analyses conducted in the last section are visualized in Figure 3. In this part of the study, within and between uniformity of LENA-provided information, MB-CDI-provided information, and CDEV-provided information were analyzed by non-parametric methods, using Spearman rank correlation coefficient ($r_s$) as an indicator of the relationship between the variables from studied measures. To enhance the reliability of the findings, all analyses were conducted using means from the pooled LENA speaker segment durations and core measures.

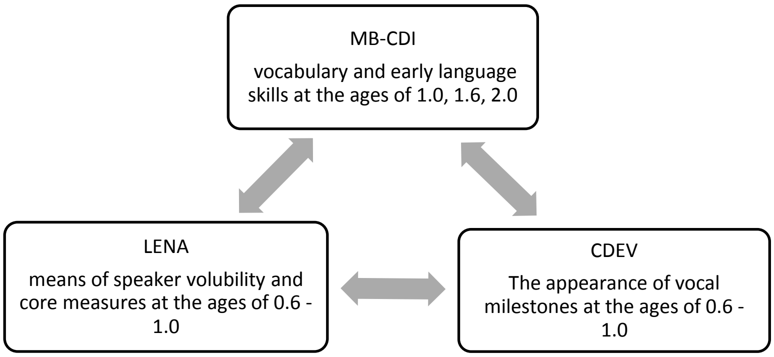



**Figure 3.** Studying the uniformity of measures and relations of language input and volubility, the appearance of milestones and vocabulary and early language skills at 1.0, 1.6 and 2.0

# 5 Results

## 5.1 LENA System reliability in this study

The first part of this study inspected the reliability of the LENA 1) speaker labelling and 2) identification of a child's speech-like vocalizations from vegetative sounds, cries, and silence. In addition, the reliability of 3) adult word and child vocalization counts were inspected. For speaker identification and vocalization detection, all the information presented in the following chapters is based on crosstabulations between LENA and human coders, which are presented in Appendix 1.

### 5.1.1 Inter-rater agreement for LENA and human speaker identification

Inter-rater agreements between human labels and machine-coding by LENA yielded moderate to high agreement rates, when "human" was set as the gold standard: Agreement was very good for key child (R1: 90%, R2: 82%) and female adult segmentation (R1: 85%, R2: 89%), moderate to good for male adults (R1: 49%, R2: 77%), and good to very good for other child (R1: 91%, R2: 70%), with overall kappa-values of $\kappa$ =.775 (R1) and $\kappa$ =.724 (R2) (Appendix 1). In addition, LENA showed good to very good sensitivity for speaker identification for all speakers (Table 7).

**Table 7.** Sensitivity, specificity, accuracy, and false discovery rates for LENA speaker identification

| | sensitivity | | specificity | | accuracy | | FDR | |
|---|---|---|---|---|---|---|---|---|
| | R1 | R2 | R1 | R2 | R1 | R2 | R1 | R2 |
| key child | .95 | .89 | .38 | .78 | .87 | .84 | .10 | .18 |
| female adult | .76 | .89 | .97 | .91 | .94 | .89 | .15 | .11 |
| male adult | .91 | .92 | .94 | .97 | .94 | .97 | .51 | .23 |
| other child | .84 | .72 | .95 | .93 | .91 | .88 | .09 | .30 |

FDR = false discovery rate

Specificity was also found to be very good with female adults, male adults, other children, and key child in R2, but not for key child in R1. A closer inspection showed that majority of misclassifications in R1 occurred in LENA between key child – other child classifications (see Appendix 1). While total accuracy of speaker identification in LENA shows to be good to very good with all speakers, false discovery rates (FDR) show a large amount of false positive errors for male adults and also a substantial amount of errors for other child in R2.

### 5.1.2 The reliability of LENA-identified key child vocalizations

The second question focused on the reliability of identification of speech-like vocalizations from other sounds present in the key child segments identified by LENA. When the human coder was used as the gold standard, overall agreement for identifying key child vocalizations yielded high agreement rates for speech-like vocalizations (R1: 84%, R2: 96%), vegetative sounds (R1: 91%, R2 85%), and for silence (R1: 96%, R2: 97%) (Appendix 1). However, agreement rates were lower for cry sounds (R1: 72% R2: 81%). For both recordings, overall agreement for key child segments was high: Cohen's Kappa values $\kappa$=.73 (R1) and $\kappa$=.83 (R2).

**Table 8.** Sensitivity, specificity, accuracy, and false discovery rates for key child segments

| | sensitivity | | specificity | | accuracy | | FDR | |
|---|---|---|---|---|---|---|---|---|
| | R1 | R2 | R1 | R2 | R1 | R2 | R1 | R2 |
| speech-like vocalization | .73 | .83 | .91 | .97 | .84 | .90 | .16 | .04 |
| vegetative sound | .57 | .72 | 1.00 | .99 | .98 | .97 | .09 | .15 |
| cry | .98 | .99 | .67 | .86 | .81 | .91 | .28 | .19 |
| silence | .95 | 1.00 | .98 | .97 | .97 | .99 | .04 | .04 |

FDR = false discovery rate

As Table 8 shows, LENA presents good sensitivity and excellent specificity for the identification of different key child sounds in this data. Sensitivity is lowest for vegetative sounds and highest for silent (pause) segments. Specificity was lowest for cry segments. Total accuracy turned out to be very good to excellent for both of the children and for all sound categories. In addition, false discovery rates for speech-like vocalizations, vegetative sounds, and pauses were found to be low, but they were higher for cry sounds.

### 5.1.3 Accuracy of LENA-provided adult word counts (AWC) and child vocalization counts (CVC)

LENA-provided daily counts differed from human counts for both measures. For AWC, LENA reported 3882 adult words from R1 in total, whereas the human count was 2559 adult words from R1. For R2, LENA counted 9431 adult words in total, while the human count was 5813 adult words, respectively. For CVC, LENA reported 762 child vocalizations for R1 in total, whereas the human counted 860 child vocalizations in total. LENA counted 782 from R2 in total, while the human count was 859 CVCs in total. As all-day counts differed substantially between machine and human, it was in the interest of the research to analyze further whether LENA errors were random or systematic. Since LENA relies on statistical probabilities, a decision was made to break up both of the recordings to recorded hours to see how AWC and CVC counts were distributed during the recording days (Figure 4).

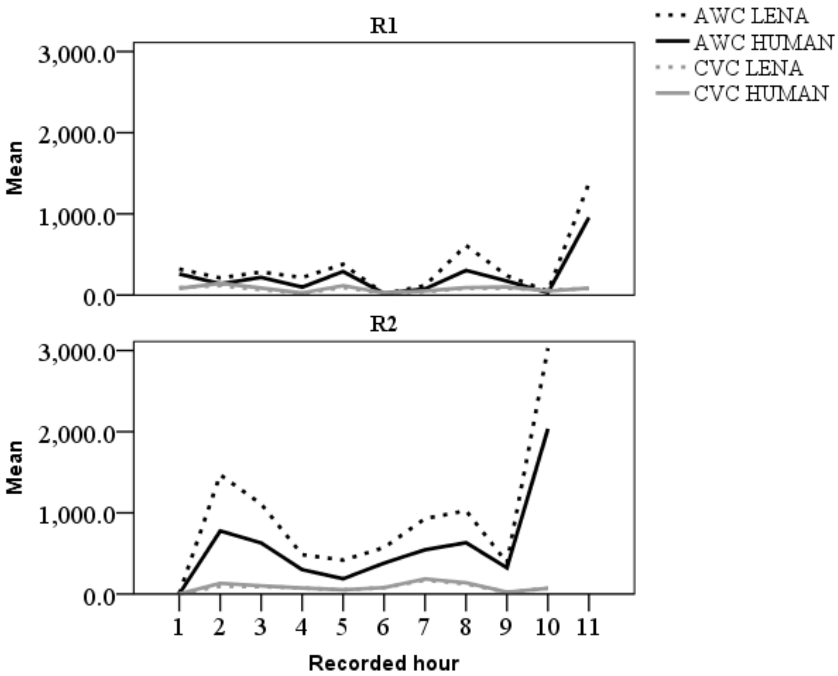



**Figure 4.** The distribution of hourly LENA and human counts in daily recordings

Machine and human CVC and AWC counts followed a similar trend, but a larger difference between LENA and human counts was found for AWC for both recordings (Figure 4). The mean difference between LENA and human coders for R1 adult words per hour was 114.8 adult words, with a median of 64.5 adult words and a 95% confidence interval for the difference from 26.5 to 203.1 adult words. The mean difference for R2 adult word counts was 361.8 adult words per hour, with a median of 301.5 adult words and a 95% confidence interval for the difference from 146.1 words to 577.5 adult words.

While LENA mainly overestimated adult words, it was found to only slightly underestimate child vocalization counts. The mean difference of R1 child vocalization counts was -8.9 counted vocalizations per hour, with a median difference of 13.1 and a 95% confidence interval for the difference from -17.7 to -0.1 vocalizations. The mean difference between coders for R2 CVC was -7.7 vocalizations per hour, with a median 11.8 and a 95% confidence interval for the difference from -16.2 to 0.8 vocalization per hour.

**Table 9.** AWC and CVC counts and relative error rates per recorded hour

| | | | AWC LENA | AWC human | AWC relative errors | CVC LENA | CVC human | CVC relative errors |
|---|---|---|---|---|---|---|---|---|
| R1 | Recorded hour | 1 | 321.0 | 258.5 | .24 | 93.0 | 82.0 | .13 |
| | | 2 | 211.0 | 139.0 | .52 | 115.0 | 145.0 | -.21 |
| | | 3 | 284.0 | 218.0 | .30 | 66.0 | 86.0 | -.23 |
| | | 4 | 216.0 | 98.0 | 1.20 | 22.0 | 27.0 | -.19 |
| | | 5 | 380.0 | 290.0 | .31 | 93.0 | 116.0 | -.20 |
| | | 6 | 15.0 | 19.0 | -.21 | 23.0 | 26.0 | -.12 |
| | | 7 | 118.0 | 76.0 | .55 | 41.0 | 51.0 | -.20 |
| | | 8 | 613.0 | 303.5 | 1.02 | 83.0 | 92.0 | -.10 |
| | | 9 | 237.0 | 172.5 | .37 | 83.0 | 99.0 | -.16 |
| | | 10 | 47.0 | 29.0 | .62 | 63.0 | 51.0 | .24 |
| | | 11 | 1380.0 | 955.5 | .44 | 80.0 | 85.0 | -.06 |
| R2 | Recorded hour | 1 | 5.0 | .0 | . | 2.0 | 2.0 | .00 |
| | | 2 | 1470.0 | 777.0 | .89 | 96.0 | 132.0 | -.27 |
| | | 3 | 1109.0 | 629.0 | .76 | 91.0 | 101.0 | -.10 |
| | | 4 | 485.0 | 300.5 | .61 | 75.0 | 74.0 | .01 |
| | | 5 | 417.0 | 188.0 | 1.22 | 51.0 | 52.0 | -.02 |
| | | 6 | 573.0 | 381.5 | .50 | 79.0 | 79.0 | .00 |
| | | 7 | 930.0 | 545.5 | .70 | 171.0 | 186.0 | -.08 |
| | | 8 | 1027.0 | 632.5 | .62 | 124.0 | 139.0 | -.11 |
| | | 9 | 386.0 | 323.0 | .20 | 21.0 | 23.0 | -.09 |
| | | 10 | 3029.0 | 2036.0 | .49 | 72.0 | 71.0 | .01 |

Table 9 presents hourly AWC and CVC counts, as well as relative error rates calculated from LENA and human counts. Error rates were not consistent throughout observed points of measure, but varied, especially for AWC in R1 (*M*= .49, min= -.21, max= 1.20, *SD*= .38) and R2 (*M*= .67, min = .20, max = 1.22, *SD* = .29). Error rate variation was more modest for CVC, with a mean relative error rate -.10 for R1 (min=-.21, max=.24, *SD*=.15) and -.06 for R2 (min=-.27, max=-.01, *SD*=.09).

### 5.1.4 Summary of key results on LENA reliability

LENA performance on speaker segment labelling was found to be adequate to very good in most of the speaker categories, including key child and female adult labelling in both of the analyzed recordings. LENA did, however, make a substantial amount of false predictions for male adult segment in R1 and for other child in R2, while error rates remained lower for male adult segments in R2 and for other child in R1.

The rate of false discoveries remained at an acceptable rate for identification of key child vocal segments from cries, vegetative sounds, and silence. LENA AWC relative error rates showed a large variation in the estimation error of adult words, when compared with the human inter-rater. However, LENA estimates on CVC were well-aligned with the interpretations of the human observer, suggesting a consistent and only a slight underestimation of child vocalizations counts.

## 5.2 Applying language environment analysis on Finnish twin data: what does LENA suggest about the language environment of Finnish twins?

In LENA, clearly spoken and speech-like segments are given the role of representatives of meaningful speech input. In twin data, clearly spoken segments made up a total of 17.1% of mean total duration of all daily recordings analyzed automatically for this study (min 14.1%, max 20.0%). This means that a 12-hour long recording included a mean of 122.4 minutes of clear human sound (including all speakers) that was detected by the program. Mean segment durations with 95 % confidence intervals of LENA-detected speaker categories are presented in Figure 5.

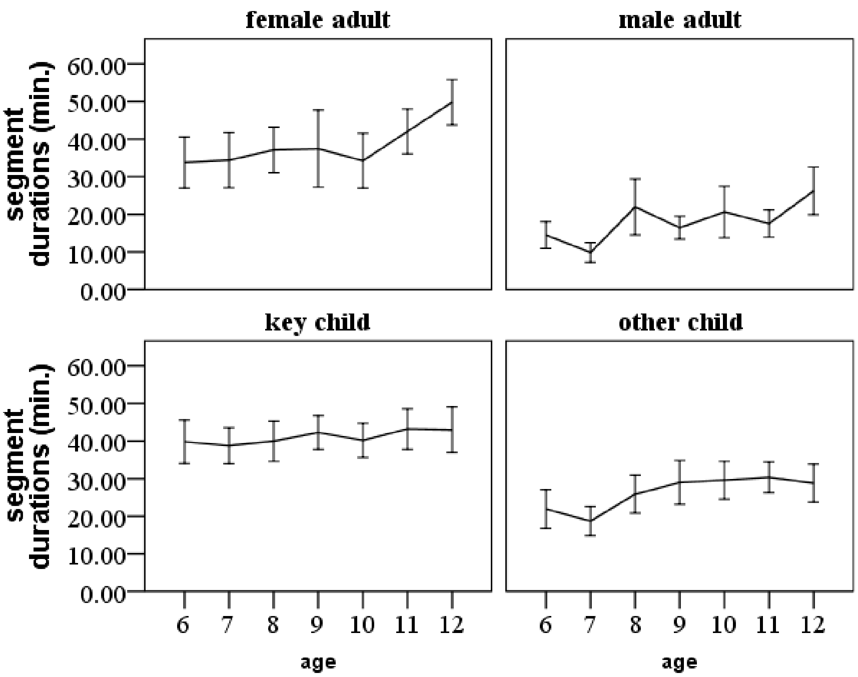



**Figure 5.** Mean speaker segment durations with 95% confidence intervals

The inspection of key child segment durations shows that volubility of the key child remained stable within the seven points of measure. In addition, the Friedman test of cross-sectional independence indicated that the distributions of variances were not statistically significantly different for key child and female adult segment durations in different points of measures, although there was a clear increase in female adult volubility from 10 months onwards. However, the variances were statistically significantly different for male adult ($F_R$=19.96, *df*=6, *p*=.003) and other child segment durations ($F_R$=42.15, *df*=6, *p*<.0001). A pairwise inspection with the adjusted significance testing revealed that the differences of variances of ranks were significant for male segments between recordings from the ages of seven and eight (*p*=.007) and seven and twelve months (*p*=.004). In other child segment durations, the variation evident in Figure 5 was confirmed to be statistically significant for the variances of ranks in earliest recordings at 6–7 months, when compared with variance ranks in the latter months of 9–12 months of age (in all *p*<.05).

### 5.2.1 Do speaker segment durations in LENA analyses differ between background variables?

#### 5.2.1.1 Key child segment duration

Group comparison with monthly inspections of key child segment durations and neonatal health-related variables did not reveal constant significant differences between any of the groups (see Appendix 2 for statistical information). In addition, no single statistically significant group difference were found for key child segment durations of birth weight, birth complications, or the level of neonatal care in any points of measure. However, preterm and full-term twins' daily segment durations were found to differ statistically significantly for key child segment durations at 10 and 11 month recordings. Preterm twins vocalized 25.2% less than full-term twins at 10 and 26.8% less than full-term twins at 11 months of CA.

When comparing key child segment durations from the pooled data of groups of twins differing in neonatal health-related background variables, no statistically significant differences were found between LBW and NBW twins, between preterm and full-term twins, or between twins born with and without birth complications (Table 10). Statistically significant differences were found in twin groups differing in the level of neonatal care; twins cared at a neonatal ward were on average 9.5% more voluble than twins cared at bedside, but the difference showed only a small practical significance.

**Table 10.** Information on central decencies, dispersion, and statistical differences in group comparisons of the effect of neonatal background variables to key child volubility

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin preterm | 42.48 | 10.74 | 40.84 | 13.41 | 55 | 2024.00 | -1.54 | 0.12 | 0.15 |
| Twin full-term | 39.96 | 11.25 | 38.16 | 13.99 | 87 | | | | |
| Twin LBW | 42.26 | 13.14 | 40.71 | 15.99 | 68 | 2246.00 | -1.10 | 0.27 | 0.11 |
| Twin NBW | 39.72 | 8.69 | 38.13 | 11.84 | 74 | | | | |
| No birth complications | 39.82 | 10.95 | 38.16 | 14.21 | 93 | 1932.00 | -1.49 | 0.14 | 0.15 |
| Born with complications | 43.05 | 11.13 | 40.84 | 11.24 | 49 | | | | |
| **Cared at bedside** | **39.35** | **9.51** | **38.03** | **13.67** | **95** | **1743.00** | **-2.12** | **0.03** | **0.22** |
| **Cared at a ward** | **44.13** | **13.27** | **42.47** | **12.25** | **47** | | | | |

No constant statistically significant group differences were found for key child segment durations in monthly group comparisons of demographic background variables, although two single statistically significant differences were detected

(Appendix 3). In addition, key child durations were not found to differ for twins differing in mother and father education, nor between twin As and Bs in any points of measure. Key child gender showed, however, a statistically significant group difference at 7 months and having an older sibling in 10-month measure points. Boy twins were on average 21.2% less voluble at seven-month recordings, when compared with twin girls and twins with siblings, who were on average 21.5% less voluble than twins without older siblings.

Table 11 presents statistical information about group comparisons of the pooled data for demographic background variables of key child segment duration. It was found that female twins' daily overall key child segment durations exceeded durations of key child male twins by 9.8%, but the effect size from the difference remained small. In addition, the mother's education did show a significant group difference with a small effect size for key child segment durations. Mothers with higher degrees had twin children who were on average 10.2% more voluble than twins of mother's with lower education. In addition, in families with no older siblings, key children were on average 46.9% more voluble than key children living in families with older siblings. However, the effect size of the mother's education on child volubility also remained small. This is explained by the proportion of overlapping variance of the observed key child durations, which is visualized in Figure 6. From other background variables, birth order and the father's education did not yield statistically significant differences in the overall daily durations of key children.

**Table 11.** Information on central decencies, dispersion, and statistical differences in group comparisons of the effect of demographic background variables to key child volubility

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin A | 41.15 | 12.09 | 40.51 | 14.71 | 71 | 2511.0 | -0.04 | 0.97 | 0.00 |
| Twin B | 40.72 | 10.06 | 38.27 | 11.44 | 71 | | | | |
| **Twin girl** | **42.58** | **11.91** | **41.88** | **12.47** | **86** | **1906.0** | **-2.10** | **0.04** | **0.21** |
| **Twin boy** | **38.41** | **9.22** | **37.83** | **10.86** | **56** | | | | |
| **No older siblings** | **73.07** | **12.03** | **40.66** | **12.88** | **71** | **2024.0** | **-2.03** | **0.04** | **0.20** |
| **Older siblings** | **38.80** | **9.67** | **7.70** | **13.56** | **71** | | | | |
| **Mom lower education** | **38.82** | **9.43** | **37.53** | **11.46** | **74** | **2012.0** | **-2.06** | **0.04** | **0.20** |
| **Mom higher education** | **43.23** | **12.30** | **40.58** | **16.26** | **68** | | | | |
| Dad lower education | 41.00 | 10.17 | 38.44 | 11.64 | 72 | 2495.0 | -1.01 | 0.31 | 0.01 |
| Dad higher education | 40.86 | 12.02 | 39.21 | 16.59 | 70 | | | | |

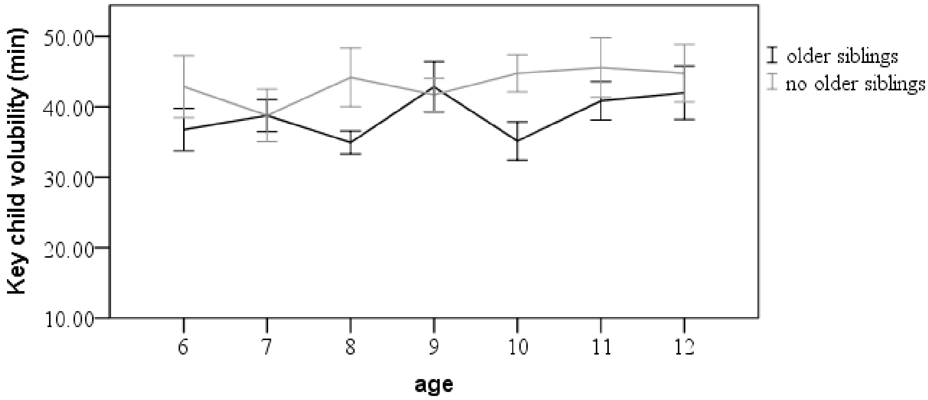



**Figure 6.** Mean key child volubility with 95% CI in families with and without older siblings

#### 5.2.1.2 Female adult segment duration

When monthly group comparisons were conducted for neonatal health-related background variables in female adult segment durations, no statistically significant group differences were detected in any points of measure between children born with and without complications (see Appendix 4 for all statistical information). In addition, no constant group differences were found between groups differing in pretermity or birth weight, although single statistically significant differences were found. Preterm children at the age of six months heard on average 44.3% less speech from female adults than full-term children, and at the age of 10 months, LBW twins heard on average 31.4% less speech from female adults than NBW children. In addition, twins cared at bedside were found to hear more female adult speech in all but one recordings, although the difference reached statistical significance only for 10- and 12-month recordings. At 10 months, children cared at bedside heard on average 49.6% more speech from female adults, when compared with children cared at a neonatal ward. However, the 12-month point of measure had markedly unequal groups due to missing values and cannot be further discussed.

When group comparisons of neonatal background variables were conducted with the pooled data, preterm and full-term twins heard a similar amount of speech from female adults to twins born with and without birth complications (see Table 12). In this data, LBW twins heard on average 17.8% less speech from female adults than NBW twins; this difference showed a medium practical significance. In addition, in families of twins who were cared at bedside during the neonatal treatment, twins heard on average 15.6% more speech from female adults than twins who had been treated at a neonatal ward. However, this difference showed small practical significance.

**Table 12.** Information on central decencies, dispersion, and statistical difference for group comparisons of the meaning of neonatal background variables to female adult volubility

| Variable groups | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin preterm | 37.51 | 15.77 | 38.43 | 27.56 | 55 | 2372.00 | -0.86 | 0.93 | 0.01 |
| Twin full-term | 38.42 | 16.47 | 38.11 | 22.61 | 87 | | | | |
| **Twin LBW** | **34.20** | **17.52** | **32.85** | **23.07** | **68** | **1745.00** | **-3.15** | **<0.001** | **0.31** |
| **Twin NBW** | **41.62** | **13.96** | **41.87** | **17.11** | **74** | | | | |
| No birth complications | 38.67 | 15.40 | 38.11 | 20 .43 | 93 | 2208.00 | -0.30 | 0.76 | 0.03 |
| Born with complications | 36.92 | 17.60 | 38.50 | 29.64 | 49 | | | | |
| **Cared at bedside** | **40.14** | **15.33** | **39.80** | **19.85** | **95** | **1665.00** | **-2.46** | **0.01** | **0.25** |
| **Cared at a ward** | **33.88** | **17.10** | **33.97** | **24.30** | **47** | | | | |

In the monthly inspection for the difference of female adult segment durations between demographic background variables, no statistically significant differences were found for groups differing in birth order, having older siblings or the father's education in any points of measure. A single statistically significant group difference was found between the female adult speech heard by girls and boys: on average, twin girls heard 42.9% less female adult speech than twin boys in the recordings obtained at 7 months. In addition, the mother's education showed a constant difference between the lower- and higher-educated mothers, although the difference reached statistical significance only in three points of measure. Children of mothers with a lower education heard female adult speech on average 40.5% more at seven, 47.5% more at nine, and 46% more at ten months of age than children of mothers with a higher education.

**Table 13.** Information on central decencies, dispersion, and statistical information for group comparisons of the meaning of demographic background variables to female adult volubility

| Variable groups | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin A | 37.66 | 16.46 | 37.26 | 25.55 | 71 | 2441.00 | -0.35 | 0.72 | 0.03 |
| Twin B | 38.48 | 15.95 | 38.50 | 22.62 | 71 | | | | |
| **Twin girl** | 36.05 | 17.47 | 36.04 | 23.50 | 86 | **1889.00** | **-2.17** | **0.03** | **0.22** |
| **Twin boy** | 41.17 | 13.45 | 42.47 | 19.47 | 56 | | | | |
| No older siblings | 37.72 | 15.90 | 38.10 | 23.28 | 71 | 2458.00 | -0.26 | 0.80 | 0.02 |
| Older siblings | 38.41 | 16.51 | 38.50 | 24.72 | 71 | | | | |
| **Mom lower education** | 44.24 | 15.43 | 44.32 | 15.46 | 74 | **1300.00** | **-4.97** | **<.0001** | **0.48** |
| **Mom higher education** | 31.35 | 14.21 | 30.78 | 17.99 | 68 | | | | |
| Dad lower education | 38.69 | 16.03 | 39.57 | 23.07 | 72 | 2282.00 | -0.16 | 0.87 | 0.09 |
| Dad higher education | 37.43 | 16.37 | 35.83 | 23.31 | 70 | | | | |

In the pooled data, boy twins heard on average 12.4% more female adult speech than girl twins, but the difference again yielded a small practical significance (Table 13). There was no statistically significant difference in female speech heard by twins for twin As and Bs. Twins with older siblings and first-born twins heard a similar amount of female adult speech, and differences in the level of the father's education did not yield significant differences in the total duration of female adult speech. However, twins, who were living in families with the mother having a lower degree, heard statistically significantly more female adult speech than twins living in families with the mother with a higher education. This difference was found to be of medium practical significance.

#### 5.2.1.3 Male adult segment duration

No statistically significant group differences were found in male adult segment durations of pretermity, birth weight, or birth complications in any points of measure in monthly inspections (see Appendix 6 for all statistical information). The only statistically significant difference for neonatal health-related background variables was found between monthly group comparisons in male adult segment durations conducted between children cared at bedside and children cared at a ward in ten-month recordings (Appendix 6). In group comparisons, children cared at bedside heard on average 58% more male adult speech than children cared at a neonatal ward.

The difference was also evident in the comparisons conducted using the pooled data: children cared at bedside heard on average 20.8% more male adult speech when compared with twins cared at a neonatal ward (Table 14). However, the difference in male adult volubility showed a small practical significance between the two groups. No other statistically significant group differences were detected in neonatal background variables related to birth weight, complications at birth, or pretermity.

**Table 14.** Information on central decencies, dispersion, and statistical difference in group differences in neonatal health-related background variables and male adult volubility

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin preterm | 17.86 | 8.92 | 19.35 | 13.35 | 55 | 2128.00 | -1.11 | 0.27 | 0.11 |
| Twin full-term | 18.01 | 13.51 | 13.73 | 15.03 | 87 | | | | |
| Twin LBW | 17.59 | 12.71 | 15.27 | 13.46 | 68 | 2345.00 | -0.70 | 0.49 | 0.07 |
| Twin NBW | 18.27 | 11.20 | 14.01 | 15.40 | 74 | | | | |
| No birth complications | 18.81 | 12.32 | 16.24 | 13.84 | 93 | 1938.00 | -1.46 | 0.14 | 0.15 |
| Born with complications | 16.30 | 11.02 | 11.83 | 14.99 | 49 | | | | |
| **Cared at bedside** | **19.27** | **12.20** | **17.37** | **15.33** | **95** | **1732.00** | **-2.13** | **0.03** | **0.22** |
| **Cared at a ward** | **15.26** | **10.92** | **13..3585** | **12.82** | **47** | | | | |

In monthly inspections of group comparisons conducted with demographic variables, no statistically significant group differences were detected in the birth order or in the mother and father's education in male adult segment durations (see Appendix 7 for statistical information). However, a constant group difference was found in mean male adult segment durations of twins with and without an older sibling, although the difference reached significance only in the measure points of 6 and 10 months. At six months, twins with older siblings heard on average 44% more male adult speech, and at ten months, on average 53% more male adult speech than twins with no older siblings.

The difference between first-born twins and twins with siblings was also evident in the pooled data. In families with older siblings, twins heard on average 28.3% more speech from male adults than in families with first-born twins (Table 15). In addition, in this data, twins living in families where the mothers had a Master's degree or a higher education heard on average 27.0% less male adult speech when compared with families where the mothers had lower educational. However, no statistically significant group differences were found at the level of the whole data in male adult volubility for twin As and Bs, for girl and boy twins, or for groups differing in the father's education.

**Table 15.** Information on central decencies, dispersion, and statistical differences in group comparisons of the effect of demographic background variables to male adult volubility

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin A | 17.48 | 11.52 | 14.40 | 14.22 | 71 | 2447.00 | -0.30 | 0.75 | 0.03 |
| Twin B | 18.41 | 12.35 | 15.28 | 14.16 | 71 | | | | |
| Twin girl | 19.37 | 13.01 | 16.61 | 13.96 | 86 | 2030.00 | -1.58 | 0.12 | 0.16 |
| Twin boy | 15.75 | 9.69 | 11.86 | 14.31 | 56 | | | | |
| **No older siblings** | **14.99** | **9.53** | **11.83** | **13.12** | **71** | **1764.00** | **-3.09** | **p<.001** | **0.30** |
| **Older siblings** | **20.90** | **13.30** | **19.13** | **14.76** | **71** | | | | |
| **Mom lower education** | **20.61** | **12.89** | **19.40** | **14.77** | **74** | **1817.00** | **-2.85** | **p<.001** | **0.28** |
| **Mom higher education** | **15.04** | **10.05** | **12.89** | **12.80** | **68** | | | | |
| Dad lower education | 17.12 | 10.33 | 13.63 | 13.92 | 72 | 2282.00 | -0.97 | 0.33 | 0.09 |
| Dad higher education | 18.79 | 13.36 | 15.94 | 14.64 | 70 | | | | |

#### 5.2.1.4 Other child segment duration

When other child segment durations were inspected in monthly group comparisons of neonatal health-related background variables, no statistically significant group differences were detected in low birth weight (LBW) and normal birth weight (NBW) children or in children born with or without complications (Appendix 8). In this data, two constant group differences were detected. Full-term children and children cared at bedside heard more other child speech or speech-like vocalizations than preterm children or children cared at a neonatal ward, although the differences reached statistical significance at six and seven months for preterm and full-term children and statistical significance at 10 months – for children cared at bedside and children cared at a ward.

**Table 16.** Information on central decencies, dispersion, and statistical differences in group comparisons of the effect of the neonatal health-related background variables to other child volubility

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| **Twin preterm** | **22.08** | **9.34** | **21.64** | **17.66** | **55** | **1657.00** | **-3.08** | **0.002** | **0.31** |
| **Twin full-term** | **28.76** | **11.41** | **27.29** | **17.08** | **87** | | | | |
| Twin LBW | 27.07 | 12.58 | 25.75 | 16.97 | 68 | 2394.00 | -0.50 | 0.618 | 0.05 |
| Twin NBW | 25.35 | 9.58 | 24.79 | 12.03 | 74 | | | | |
| No birth complications | 27.32 | 11.81 | 25.92 | 15.77 | 93 | 1977.00 | -1.29 | 0.196 | 0.13 |
| Born with complications | 24.00 | 9.37 | 22.89 | 13.96 | 49 | | | | |
| **Cared at bedside** | **27.99** | **10.97** | **27.53** | **14.35** | **95** | **1550.00** | **-2.96** | **0.003** | **0.31** |
| **Cared at a ward** | **22.51** | **10.58** | **19.95** | **12.73** | **47** | | | | |

When group comparisons were conducted with pooled data, two neonatal health-related group comparisons reached statistically significant differences. Preterm twins heard on average 23.2% less other child speech or speech-like vocalizations than twins born full-term; this was shown to have medium practical significance (Table 16). No statistically significant group differences were found for other child volubility between LBW and NBW twins or twins born with or without birth complications.

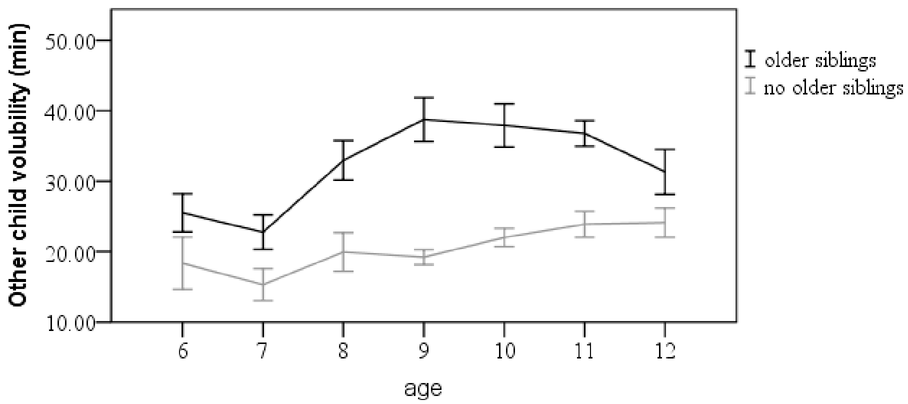



**Figure 7.** Mean segment durations and 95% confidence intervals of other child segments in families with and without older siblings

When group comparisons were conducted for demographic background variables within monthly inspections, no statistically significant group differences were detected in birth order, gender, or parental education in any points of measure (Appendix 9). However, a constant group difference with statistically significant differences in five out of seven points of measure was found for other child volubility in families with and without older siblings (see Figure 7).

In comparisons conducted with demographic background variables with the pooled data, no statistically significant group differences were detected in maternal and paternal education, nor for twins' birth order. However, statistically significant group differences were found in other child segment durations between girl and boy twins and between twins with and without siblings. In the whole data, twin girls heard on average 20.8% more other child speech or speech-like vocalizations than boy twins; the difference was found to be statistically significant with small to medium practical importance (Table 17). In addition, twins with older siblings heard on average 37.8% more of other child speech or speech-like vocalizations when compared with first-born twins. This difference showed a large practical significance.

**Table 17.** Information of central decencies, dispersion and statistical for group comparisons of the effect of demographic background variables to other child volubility

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin A | 25.99 | 11.91 | 23.08 | 15.57 | 71 | | | | |
| Twin B | 26.35 | 10.33 | 25.68 | 14.59 | 71 | 2405.00 | -0.47 | 0.637 | 0.05 |
| **Twin girl** | **28.51** | **11.65** | **27.41** | **15.96** | **86** | | | | |
| **Twin boy** | **22.59** | **9.23** | **21.96** | **13.28** | **56** | **1707.00** | **-2.93** | **0.003** | **0.29** |
| **No older siblings** | **20.07** | **7.97** | **19.05** | **9.46** | **71** | | | | |
| **Older siblings** | **32.28** | **10.46** | **31.72** | **12.25** | **71** | **830.00** | **6.90** | **<.0001** | **0.67** |
| Mom lower education | 25.52 | 10.13 | 25.31 | 13.31 | 74 | | | | |
| Mom higher education | 26.89 | 12.13 | 25.28 | 17.53 | 68 | 2394.00 | -0.50 | 0.618 | 0.05 |
| Dad lower education | 26.32 | 9.08 | 26.59 | 12.87 | 72 | | | | |
| Dad higher education | 26.02 | 12.94 | 23.06 | 17.73 | 70 | 2272.00 | -1.01 | 0.31 | 0.10 |

## 5.2.2 Measuring twin family interaction with LENA Core measures

### 5.2.2.1 LENA AWC estimates

All LENA-provided mean daily AWC estimates from twin data fitted to LENA norms are presented in Figure 8. Contrary to stable LENA AWC norms, the amount of adult words in twin data showed a rising trend with modest positive correlation ($r_S$=.38, *p*<.001) of AWC increase with the increasing age of the children. In addition, the variances of AWCs were found to be significantly different in different points of measure ($F_R$=31.57, *df*=6, *p*<.0001). A pairwise inspection with adjusted significance testing revealed that the differences of the variances of ranks were significant for LENA estimates of AWC in the counts from 12 months of age and six (*p*=.001), seven (*p*<.0001), and nine months of age.

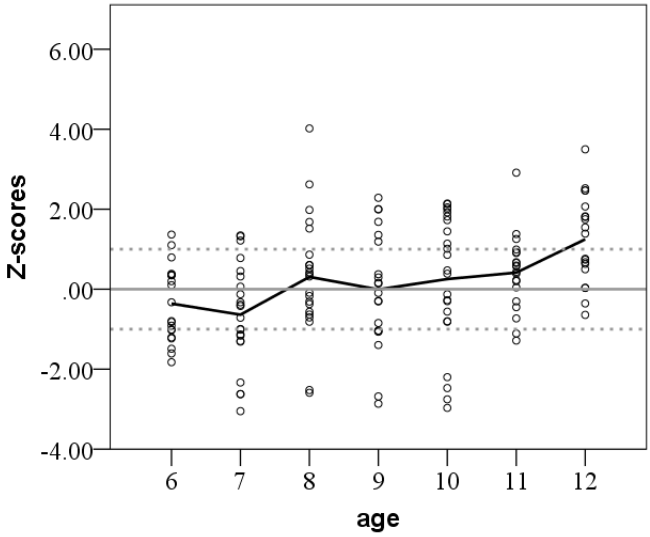



**Figure 8.** AWC from twin data, adjusted to LENA norms

No constant group differences were detected in monthly comparisons of children born with and without birth complications (see statistical information in Appendix 10). However, LENA AWC estimates did statistically significantly differ for preterm and full-term children at six months of age and for LBW and NBW twins at 10 months of age. According to LENA, preterm twins heard 28.9% more adult words at six months of age when compared with full-term twins, and NBW twins heard 55% more adult words at ten months of age when compared with LBW twins. In addition, significant differences were also detected in groups differing in the level of neonatal care at the ages of 10 and 12 months. At 10 months, twins who had been cared at bedside heard 50.0% more adult words according to LENA estimates than children who were cared at a neonatal ward. However, the 12-month difference was not inspected closely because of unequal groups due to a few missing recordings (bedside n=14, ward n=3, see Appendix 11).

When the group comparisons were conducted with neonatal background variables in the pooled data, no statistically significant group differences in the amount of LENA-calculated AWC were found in families of preterm and full-term twins or twins born with and without birth complications (see Table 18). Mean group differences of LENA-estimated word counts from the ages of 6 to 12 months were found to be statistically significantly different for birth weight and the level of neonatal care, although with small practical difference. In the whole data, mean calculations from LENA estimates suggested that NBW twins heard on average 12.8% more words during recordings than LBW twins. In addition, according to LENA estimates, twins cared at bedside heard on average 16.7% more words during recording days than twins who were treated at a neonatal ward after birth.

**Table 18.** Information on central decencies, dispersion, and statistical differences in group comparisons of the effect of the neonatal health-related background variables on AWC

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Preterm | 13579.78 | 5274.89 | 12893.28 | 9589.16 | 55 | 2281.0 | -0.47 | 0.64 | 0.05 |
| Full-term | 13216.27 | 6190.19 | 13273.37 | 7296.85 | 87 | | | | |
| **LBW** | **12403.56** | **6548.95** | **11561.08** | **7353.84** | **68** | **1931.0** | **-2.39** | **0.02** | **0.23** |
| **NBW** | **14223.26** | **4979.66** | **13743.67** | **7713.01** | **74** | | | | |
| No complications | 13658.11 | 5530.73 | 13273.37 | 6621.40 | 93 | 2074.0 | -0.88 | 0.38 | 0.09 |
| Complications | 12786.70 | 6396.20 | 12991.62 | 10586.24 | 49 | | | | |
| **Care at bedside** | **14128.12** | **5400.78** | **13616.64** | **7623.14** | **95** | **1632.0** | **-2.60** | **0.01** | **0.27** |
| **Care at a ward** | **11798.54** | **6409.72** | **10794.46** | **7960.80** | **47** | | | | |

When monthly group comparisons were conducted with demographic background variables, no group differences were found in any points of measure in groups differing in birth order, having older siblings, or in the father's educational level (Appendix 11). However, LENA-provided AWC estimates did differ statistically significantly in twins in families differing in the mother's education at seven and nine-month recordings. In families with the mother having a lower education, children both at seven and nine months of age heard 43% more adult words, according to LENA AWC estimates, than children from families, where mothers had a higher educational level. Additionally, the amount of AWC estimated by LENA differed statistically significantly for boy and girl twins at seven-month recordings. According to LENA, boy twins heard 31.1% more adult words than girl twins at seven months of age.

When group comparisons were conducted with the pooled data, no significant differences between demographic background variables were found, when comparing AWC in recordings from boy and girl twins, twin As and Bs, twins with and without older siblings, and recordings from families differing in the father's education (Table 19). The only background variable with group difference on the whole data was found in families differing in maternal education. In families where mothers had a lower educational level, adults were found to produce on average 30.1% more words according to LENA estimates, when compared with families of mothers with higher educations. This difference showed medium practical significance.

**Table 19.** Information on central decencies, dispersion, and statistical difference in group comparisons of the effect of demographic background variables to AWC

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin A | 13032.76 | 5701.32 | 12639.55 | 8056.31 | 71 | 2383.0 | -0.56 | 0.58 | 0.05 |
| Twin B | 13681.37 | 5990.56 | 13370.56 | 8636.72 | 71 | | | | |
| Twin girl | 13328.14 | 6423.69 | 12805.77 | 7261.79 | 86 | 2316.0 | -0.38 | 0.70 | 0.04 |
| Twin boy | 13401.49 | 4851.90 | 13371.77 | 8443.81 | 56 | | | | |
| No older siblings | 12525.33 | 5637.73 | 12722.55 | 7744.91 | 71 | 2121.0 | -1.63 | 0.10 | 0.16 |
| Older siblings | 14188.80 | 5951.29 | 13569.70 | 8211.43 | 71 | | | | |
| **Mom lower education** | **15606.48** | **5660.81** | **14878.34** | **8872.99** | **74** | **1353.0** | **-4.75** | **p<.0001** | **0.46** |
| **Mom higher education** | **10909.17** | **5009.23** | **10737.98** | **6099.97** | **68** | | | | |
| Dad lower education | 13255.64 | 5559.56 | 13421.54 | 9056.04 | 72 | 2499.0 | -0.08 | 0.932 | 0.01 |
| Dad higher education | 13461.39 | 6145.98 | 12891.13 | 6920.22 | 70 | | | | |

### 5.2.2.2 LENA CVC estimates

On average, CVC counts suggested that children in twin data seemed to vocalize slightly more than children in US English normative data (Figure 9). No statistically significant correlations were found for norm-adjusted CVC and age ($r_s$=-.09. *p*=.29). In addition, the Friedman test of cross-sectional independence indicated that the distributions of variances were not statistically significantly different for CVCs in different points of measure (*p*=.27).

Monthly group comparisons of neonatal health-related background variables did not reveal any constant group differences in any of the groups, and no statistically significant group differences were detected in CVC between twins cared at bedside and twins cared at a neonatal ward (see statistical information on Appendix 12). However, single statistically significant group differences were detected for CVC between children born with and without birth complications at seven months, between preterm and full-term twins at 11 months, and between LBW and NBW twins at 12 months. LENA estimated that children born with birth complications vocalized on average 24.4% more than children born without birth complications, that preterm children vocalized on average

26.3% more than full-term children, and that children with LBW vocalized 30.6% less than NBW children.

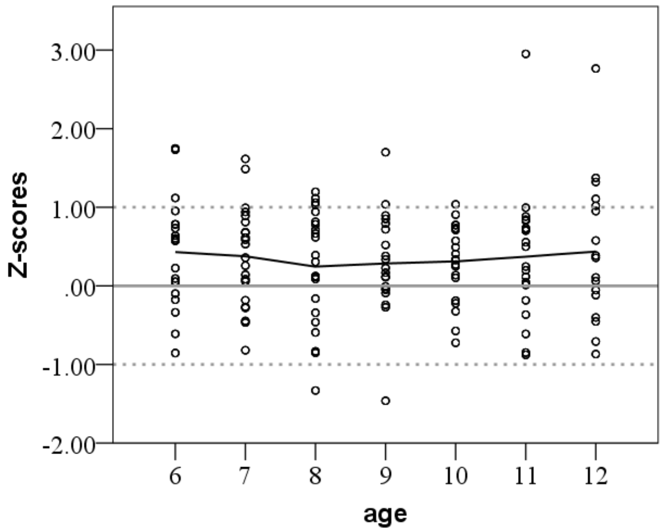



**Figure 9.** Twin CVC adjusted to LENA norms

However, when CVC estimates from the pooled data were compared in groups differing in neonatal health-related variables, no statistically significant group differences were detected between preterm and full-term twins, between twins born with and without complications, nor between twins cared at bedside and twins cared at a neonatal ward (Table 20). A statistically significant group difference with small practical significance was discovered in the amount of CVC between LBW and NBW groups. LENA estimated that during the 6-12 months recordings, LBW children vocalized on average 11% less than NBW children.

**Table 20.** Information on central decencies, dispersion, and statistical differences in group comparisons of the effect of the neonatal health-related background variables to CVC

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin preterm | 1266.20 | 443.16 | 1255.19 | 470.86 | 55 | 2381.00 | -0.05 | 0.96 | 0.00 |
| Twin full-term | 1252.96 | 381.77 | 1236.93 | 509.65 | 87 | | | | |
| **Twin LBW** | **1188.29** | **434.20** | **1166.37** | **593.26** | **68** | **1942.00** | **-2.34** | **0.02** | **0.23** |
| **Twin NBW** | **1322.23** | **367.86** | **1273.85** | **402.99** | **74** | | | | |
| No complications | 1232.75 | 446.24 | 1174.19 | 536.36 | 93 | 1876.00 | -1.73 | 0.08 | 0.18 |
| Born with complications | 1306.19 | 311.34 | 1272.75 | 348.89 | 49 | | | | |
| Cared at bedside | 1265.37 | 383.11 | 1248.50 | 474.94 | 95 | 2111.00 | -0.53 | 0.60 | 0.05 |
| Cared at a ward | 1243.38 | 450.46 | 1208.82 | 514.00 | 47 | | | | |

In monthly group comparisons of CVCs, no constant statistically significant differences were found between any of the demographic background variables (Appendix 13). There were also no statistically significant group differences in CVC in any of the points of measure for twin As and Bs, girls and boys, and for twins from families differing in maternal education. Single statistically significant differences for CVC were found for twins differing in the father's education at seven months and for having older siblings at ten months. In LENA CVC estimates, children whose fathers were in the group of lower education were found to vocalize on average 34.0% more than children who had fathers with a higher education. In addition, first-born twins vocalized at 12 months on average 24.6% more than twins with older siblings; this difference showed very large practical significance. However, neither of these groups had other statistically significant differences in any other points of measure.

When group comparisons were conducted from the pooled data, statistically significant group differences with small practical significance were found for CVC between twin girls and boys and between twins with and without older siblings (Table 21). In the whole data, LENA estimations of CVCs were greater for twin boys than for twin girls and lower for twins with older siblings, when compared with first-born twins. According to LENA CVC, girls vocalized on average 9.1% less than twin boys within the 6-12 months of measures, and twins with older siblings vocalized on average 10.3% less than first-born twins. Statistically significant group differences were not detected in A and B-twins, nor in groups differing in parental education.

**Table 21.** Information on central decencies, dispersion, and statistical difference in group comparisons of the effect of demographic background variables to CVC

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin A | 1273.38 | 416.90 | 1253.33 | 531.97 | 71 | 2390.00 | -0.53 | 0.59 | 0.05 |
| Twin B | 1242.81 | 395.43 | 1200.74 | 444.51 | 71 | | | | |
| **Twin girl** | **1211.62** | **398.31** | **1173.90** | **501.59** | **86** | **1874.00** | **-2.23** | **0.03** | **0.22** |
| **Twin boy** | **1329.45** | **408.76** | **1348.18** | **447.56** | **56** | | | | |
| **No older siblings** | **1190.04** | **336.31** | **1168.51** | **468.32** | **71** | **2048.00** | **-1.93** | **0.05** | **0.19** |
| **Older siblings** | **1326.14** | **456.23** | **1283.57** | **520.77** | **71** | | | | |
| Mom lower education | 1236.48 | 407.91 | 1199.66 | 533.39 | 74 | 2293.00 | -0.91 | 0.36 | 0.09 |
| Mom higher education | 1281.61 | 403.84 | 1263.31 | 497.92 | 68 | | | | |
| Dad lower education | 1300.21 | 280.54 | 1264.85 | 317.05 | 72 | 2499.00 | -0.09 | 0.93 | 0.01 |
| Dad higher education | 1214.76 | 500.71 | 1115.40 | 661.17 | 70 | | | | |

#### 5.2.2.3 LENA CTC estimates

Figure 10 presents the distribution of norm-adjusted CTC from twin data. CTC was found to be moderately correlated with twin children's increasing age ($r_s$=.41), thus suggesting that twin families in this data became more active in engaging in turn-taking with infants as twins got older. In addition, the variances of CTCs were found to be significantly different in different points of measure ($F_R$=30.66, *df*=6, *p*<.0001). A closer inspection using the adjusted significance testing revealed that the differences of the variances of ranks were significant for LENA estimates of CTC between counts from 6, 7, 8, and 9 months and counts from 12 months of age (all *p*≤.001), but not significant between 10 and 11 months to 12 months.

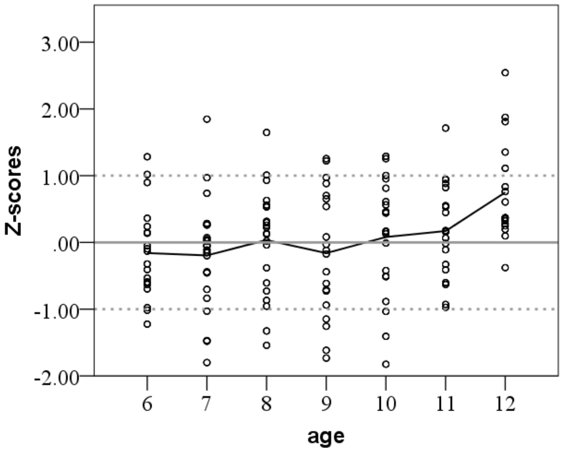



**Figure 10.** The distribution of twin family CTC in comparison to LENA norms

In the monthly group comparison of LENA CTC estimates of neonatal health-related background variables, no statistically significant differences were found between children who were born with and without birth complications. Single statistically significant differences were found between preterm and full-term and between children cared at bedside and children cared at a neonatal ward (see Appendix 14 for statistical information). According to LENA, preterm children participated 35% less in conversational turns than full-term children at ten months, and children treated at a ward participated in conversational turns on average 38.8% less often than children cared at bedside. The most constant variable with significant group differences in CTC estimates was found for birth weight. In all measure points LBW children participated in fewer turns than NBW children, but the difference reached statistically significant difference at 7, 9, and 10 months. LBW children participated in conversational turns on average 34.3% less at seven, 41.1% less at nine, and 41.6% less at ten months of age.

When group comparisons were conducted from the pooled data, neonatal health-related group comparisons showed statistically significant differences with medium to large practical significance for CTCs between LBW and NBW children and between children cared at bedside and children cared at a neonatal ward (Table 22). According to LENA estimates, NBW children were involved in conversations on average 32.5% more often than LBW children during the follow-up from six to twelve months, and children who were cared at bedside were involved in conversational turns on average 25.8% more often than children who had been cared at a neonatal ward.

**Table 22.** Information on central decencies, dispersion, and statistical differences in group comparisons of the effect of the neonatal health-related background variables to CTC

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin preterm | 317.15 | 135.63 | 300.32 | 202.16 | 55 | 2135.00 | -1.08 | 0.281 | 0.11 |
| Twin full-term | 292.74 | 141.47 | 277.75 | 175.54 | 87 | | | | |
| **Twin LBW** | **241.49** | **108.46** | **227.61** | **133.34** | **68** | **1217.00** | **-5.35** | **p<.0001** | **0.52** |
| **Twin NBW** | **357.98** | **141.69** | **363.78** | **185.93** | **74** | | | | |
| No birth complications | 304.72 | 136.44 | 294.00 | 175.14 | 93 | 2208.00 | -0.30 | 0.762 | 0.03 |
| Born with complications | 297.40 | 145.78 | 274.38 | 219.82 | 49 | | | | |
| **Cared at bedside** | **330.36** | **141.34** | **317.44** | **181.49** | **95** | **1372.00** | **-3.73** | **p<.0001** | **0.39** |
| **Cared at a ward** | **245.26** | **116.93** | **228.86** | **134.15** | **47** | | | | |

In monthly group comparisons of CTC estimates with demographic background variables, no statistically significant group comparisons were detected between twin As and Bs, between twins with and without older siblings, or groups differing in the father's education in any points of measure (Appendix 15). Groups differing in the mother's education yielded statistically significant differences in two measure points: twins from families with a lower-educated mother were engaged in conversational turns on average 38.1% more often at the age of seven months and on average 31% more often at the age of 10 months. The most constant difference was found in LENA CTC estimates between twin girls and boys. Girls were engaged in fewer conversations in all but one recording; these differences were statistically significantly different in three points of measure. Girls were engaged in conversational turns on average 39.9% less often at seven months, 39.4% at nine, and 28.6% less often at ten months of age.

When group comparisons were conducted from the pooled data for demographic background variables, statistically significant group differences with medium practical significance were detected in CTC estimates in groups differing in twin gender and mothers' educations (Table 23). According to LENA, twin girls participated in conversational turns between the 6-12 months on average 22.3% less than twin boys.

Additionally, participation in conversational turns remained on average 24.4% lower for twins from families with the mother having a higher education, when compared with twins from families where the mother had lower education. No statistically significant differences were found for LENA CTC estimates of the group comparisons for the whole data between A-twins and B-twins, twins with and without older siblings, nor twins differing in the father's education.

**Table 23.** Information on central decencies, dispersion, and statistical difference for group comparisons of the effect of demographic background variables to CTC

| Variable group | *M* | *SD* | *Mdn* | *IQR* | N | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|
| Twin A | 297.06 | 128.69 | 295.98 | 169.23 | 71 | 2490.00 | -0.12 | 0.901 | 0.01 |
| Twin B | 307.33 | 149.84 | 285.34 | 202.45 | 71 | | | | |
| **Twin girl** | **271.44** | **129.26** | **254.53** | **187.85** | **86** | **1587.00** | **-3.43** | **0.001** | **0.34** |
| **Twin boy** | **349.42** | **141.95** | **352.59** | **183.88** | **56** | | | | |
| No older siblings | 306.52 | 152.95 | 286.87 | 199.32 | 71 | 2520.00 | 0.00 | 0.998 | 0.00 |
| Older siblings | 297.87 | 125.04 | 294.01 | 173.32 | 71 | | | | |
| **Mom lower education** | **342.18** | **144.32** | **323.41** | **197.06** | **74** | **1645.00** | **-3.56** | **p<.0001** | **0.35** |
| **Mom higher education** | **258.69** | **120.15** | **259.88** | **181.47** | **68** | | | | |
| Dad lower education | 317.50 | 133.94 | 315.29 | 202.03 | 72 | 2080.00 | -1.80 | 0.073 | 0.17 |
| Dad higher education | 286.45 | 143.79 | 267.38 | 190.22 | 70 | | | | |

### 5.2.3 Summary of key results of automated measurement of twins' language environment

LENA-detected speech and speech-like vocalizations accounted for 17% of all recorded audio. In monthly inspections, it was shown that key child volubility remained stable over time, but there was an increase in female adult, male adult, and other child speaker categories. When key child segment durations were compared between background variables, the difference between children cared at bedside and children cared at a neonatal ward was present in six out of seven recordings, but the difference remained small in the pooled data. In addition, twins with older siblings were less voluble than first-born twins, girls were found to be slightly more voluble than boys, and children from higher-educated mothers were slightly more voluble than children from lower-educated mothers. No other statistically significant differences were detected in group comparisons.

The inspection of the mean female adult segment durations showed that in six of the seven monthly measure points, children cared at bedside heard more female adult speech than children cared at a neonatal ward. This difference was also evident in the comparison with the pooled data, although showing small practical significance. In addition, twins of low birth weight were found to hear less female adult speech than twins of normal birth weight, when comparisons were conducted using the pooled data. This difference showed medium practical significance. In addition, mothers with a lower education talked substantially more than mothers with a higher education; this difference was evident both in monthly inspections and in the group comparisons conducted using pooled data. In addition, twin boys heard slightly more female adult talk than twin girls in this data. However, this difference showed only small practical significance in the comparisons conducted using the pooled data. No other statistically significant group differences were detected in group comparisons.

The group comparisons of segment durations of male adults showed that children cared at bedside heard more male adult speech than children cared at a neonatal ward. In addition, the amount of male adult speech was found to be higher in families with lower-educated mothers. Besides mothers' education, having older siblings showed medium practical significance to mean male adult segment durations. Male adults were found to talk more in families with older siblings. Both of the differences showed small to medium practical significance. No other statistically significant group differences were detected in group comparisons.

Constant group differences were detected for other child segment durations between preterm and full-term children and between children cared at bedside and children cared at a neonatal ward. In the pooled data, pretermity and neonatal care showed medium practical significance for other child volubility. Full-term twins and twins cared at bedside heard more other child speech or speech-like vocalizations than preterm twins and twins cared at a neonatal ward. In addition, in this data, female twins heard more other child speech than twin boys, a difference showing only small practical significance in the pooled data. However, children with older siblings heard a substantial amount more of other child speech than children without older siblings; this difference showed large practical importance.

Contrary to stable LENA norms of the core measures, adult words and conversational turns were found to increase with age in twin families, but no increase was detected for child vocalizations. No constant group differences were detected for any of the background variables in child vocalization and adult word counts, but differences in birth weight and gender showed constant group differences in monthly comparisons in conversational turn count. In the pooled data, however, according to LENA, children of low birth weight and children cared at a neonatal ward were found to hear slightly less adult words, when compared with children of normal birth weight and children cared at bedside. In addition, twins of low birth weight, girls, and twins with older siblings vocalized slightly less on average than twins of normal birth weight, boys, and first-born twins. Mothers with a lower education were found to produce more words than mothers with a higher education; this difference showed medium to large practical significance. Conversational turns were more frequent in families when the mother had a lower education, with children of normal birth weight, and cared at bedside, when compared with the amount of conversations in families with the mother having a higher education or with conversational turn count of low birth weight children and children cared at a neonatal ward.

## 5.3 Milestones, vocabulary and early language skills

### 5.3.1 Appearance of vocal milestones

Twins began reduplicative babbling in the mean age of 31 weeks of corrected age (*Mdn*=31, *SD*=8.7), which falls three weeks behind from the mean presented in the normative data of the CDEV (*M*=29, *Mdn*=27, *SD*=3.8). No significant group differences were found in birth weight, pretermity, complications at birth, the level of neonatal care, gender, having older siblings, or parental education to affect the onset of reduplicated babbling (see Appendix 16 for all statistical information). However, the distribution of the onset of reduplicated babbling differed statistically significantly in categories of birth order. A-children began reduplicative babbling four weeks earlier (*M*=29.0, *Mdn*=29.5, *SD*=3.0) than B-children (*M*=33.0, *Mdn*=33.0, *SD*=3.7), a difference showing substantial practical significance (*U*=19.000, *Z*=-2.367, *p*=0.19, $r_{rb}$=.62). No other statistically significant group differences were detected between any other background variables.

The onset of variegated babbling was discovered in 12 of the twins before the onset of words. In eight out of ten pairs, at least one of the twins started variegated babbling before first words, and out of these, only one B-twin was reported to start variegated babbling before A-twin. The mean age of the onset of variegated babbling of the 12 twins was 46 weeks of corrected age in parent reports (*Mdn*=44, *SD*=12.4), which falls 10 weeks behind the mean presented in checklist normative data (*M*=36, *Mdn*=32, *SD*=9.1) (Lyytinen *et al.*, 1996). As nearly half of the twins did not acquire variegated babbling before the onset of first words, no statistical hypothesis testing was conducted on group differences between any of the background variables.

The mean emergence of protowords from parent reports was 47 weeks of CA (*Mdn*=48, *SD*=6.7), but the information cannot be compared to normative information presented in the method manual due to missing statistical information from the method manual. No statistically significant differences were found in group comparisons of birth weight, pretermity, complications at birth, the level of neonatal care, birth order, gender, siblingship, or the father's education. However, a statistically significant group difference was detected in groups differing in maternal education and birth weight. Children of mothers with a higher education were seen to produce protowords on average nine weeks earlier (*M*=41.1, *Mdn*=44.0, *SD*=6.9) than children from mothers with a lower education (*M*=50.5, *Mdn*=50, *SD*=4.5) – a difference that shows large practical significance (*U*=5.5, *Z*=-2.9, *p*=.002, $r_{rb}$ =.84). In addition, the emergence of protowords was discovered earlier in LBW twins (*M*=42.6, *Mdn*=46.0, SD=7.7) when compared with NBW twins (*M*=49.5, *Mdn*=48.5, *SD*=5.5); this difference showed a substantial practical significance (*U*=24.5, *Z*= -2.0, *p*=.04. $r_{rb}$=.59).

### 5.3.2 Receptive vocabularies at 12 and productive vocabularies at 12, 18, and 24 months of corrected age

Twin's mean receptive vocabulary fell short of the normative data reported in Lyytinen (1999). in the present study, twins had a mean of 56.4 receptive words at 12 months of corrected age (*Mdn*=43, *SD*=42), while children in the norm study had a mean of 89.3 words in their receptive vocabularies (*Mdn*=67, *SD*=63.2) (see Figure 11). In the group comparisons, a statistically significant difference in receptive vocabularies was found between twins born with and without complications. Children born without complications had significantly larger (*M*=70.2, *Mdn*=76.0, *SD*=41.7) receptive vocabularies at 12 months, when compared with children born through complicated births (*M*=29.7, *Mdn*=27, *SD*=31.3). This difference showed medium practical significance (*U*=19.5, *Z*=-2.1, *p*=.037, $r_{rb}$=.57). No other statistically significant group differences were detected for 12-month receptive vocabulary (see all statistical information on vocabulary comparisons for 12, 18, and 24 months in Appendix 17).

At the age of 12 months, the mean productive vocabulary of twins was 2.5 words (*Mdn*=2.5, *SD*=2.5), which is smaller than the mean reported in the normative study (*M*=7.1, *Mdn*=3.5, *SD*=9.44). In the group comparisons, it was found that LBW twin children produced more words (*M*=4.1, *Mdn*=4.0, *SD*=2.6) than NBW children (*M*=1.6, *Mdn*=1.0, *SD*=1.8), and that the difference showed a substantial practical significance (*U*=19.5, *Z*=-2.35, *p*=.019, $r_{rb}$ =.61). In addition, statistically significant group differences were detected for groups differing in the father's education both in receptive and expressive vocabulary. Children of lower-educated fathers understood a mean of 32.3 words at 12 months (*Mdn*=31.5, *SD*=27.9), while children with a higher-educated father understood a mean of 79.7 words at 12 months (*Mdn*=85.5, *SD*=42.1) – a difference showing substantial practical significance (*U*=18.0, *Z*=-2.42, *p*=.015, $r_{rb}$=.64). In addition, children with higher-educated fathers knew more words at 12 months (*M*=4.3, *Mdn*=4.0, *SD*=2.5), when compared with children with lower-educated fathers (*M*=1.4, *Mdn*=1.0, *SD*=1.6);this difference also showed a substantial practical significance (*U*=15.5, *Z*=-2.66, *p*=.007, $r_{rb}$ =.69). No other statistically significant differences were found in group comparisons in any demographic background variables in 12-month receptive and expressive development.

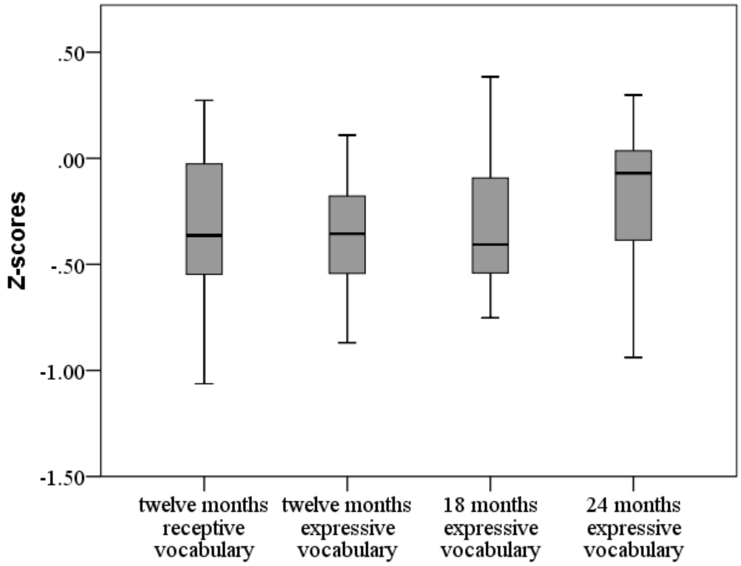



**Figure 11.** Twin's vocabularies adjusted to Finnish MB-CDI norms

As MB-CDI does not differentiate receptive and expressive vocabularies from 18 months onwards, the studies from vocabulary development at the ages of 18 and 24 months focus solely on the development of expressive vocabularies. At 18 months of corrected age, the mean productive vocabulary was 37.3 words for twins, with a median of 23 and a standard deviation of 34.8 words. Thus, twins' mean vocabulary counts continued to be lower than reported in norms. Lyytinen (1999) reported the mean productive vocabulary to be 70.4, with a median of 34 and the standard deviation of 78.4. In group comparisons, a statistically significant difference was found in vocabularies between term and preterm twins. Preterm twins had a mean of 38.1 words (*SD*=24.4, *Mdn*=36.5), while term twins had a mean of 36.8 words (*SD*=41.3, *Mdn*=18). This difference showed substantial practical significance (*U*=17.0, *Z*=-2.54, *p*=.016, $r_{rb}$ =.65). No other statistically significant differences were found for 18-month vocabulary between any other groups.

At 24 months, the mean productive vocabulary for twins was 231.83 words (*Mdn*=249, *SD*=104.7), which remained lower than the reported normative mean presented in the method manual (*M*=277.9, *Mdn*=269, *SD*=162.7). A statistically significant difference in favor of twins with older siblings was discovered in group comparisons. While the mean word count for first-born twins was 161 words, with a median of 129.5 words (*SD*=93.7), twins with older siblings had substantially larger productive vocabularies with a mean 288.5 words and a median of 291.5 words (*SD*=76.5) at 24 months. This group difference showed substantial practical significance (*U*=17.0, *Z*=-2.5, *p*=.011$r_{rb}$=.66). No other statistically significant group differences were detected between background variables in the amount of vocabulary at 24 months.

### 5.3.3 Language skills at 12, 18, and 24 months

Children's abilities to understand language and use non-verbal means of communication were measured with MB-CDI First signs of understanding and the sum variable of actions and gestures. On average, parents of twins considered their children to show <u>early signs of comprehension</u> at the age of 12 months (*M*=2.8, *Mdn*=3.0) similarly to Finnish parents previously reporting in the MB-CDI norm study (Lyytinen, 1999). No significant differences were found in group comparisons conducted with background variables of birth weight, siblings, pretermity, the level of neonatal care, and gender. However, children born with complications at birth scored lower on early signs of comprehension in this data (with *M*=2.4, *SD*=0.5, *Mdn*=2.0) than children born with no birth complications (*M*=3, *SD*=0, *Mdn*=3); this difference showed medium practical significance (*U*=19.5, *Z*=-2.97, *p*=.037, $r_{rb}$=. 57). (see all statistical information about 12-month group comparisons in Appendix 18).

The mean of the sum variable score of <u>actions and gestures</u> of twins at the age of 12 months was 28.5 (*Mdn*=28, *SD*=5.0), falling close to the standard mean presented in the method manual. In addition, the distribution of sum variable scores in twin data and their standard deviation fell within the standard deviation and range presented in the method manual (*M*=29.4, *Mdn*=29.0, *SD*=8.1). In group comparisons, statistically significant group differences were not detected in any of the neonatal health-related background variables, nor in birth order, gender, or the father's education. However, statistically significant group differences were found between children with and without siblings and between children from families differing in the mother's education. Children with siblings scored higher on non-verbal communicative actions and gestures (*M*=31.5, *SD*=4.7, *Mdn*=31), when compared with first-born twins (*M*=25.5, *SD*=3.2, *Mdn*=26.5) – a difference showing large practical significance (*U*=8.0, *Z*=-2.53, p=.01, $r_{rb}$=.84). Additionally, children of lower-educated mothers scored higher in actions and gestures (*M*=31.1, *SD*=3.9, *Mdn*=31), when compared with children from higher-educated mothers (*M*=24.2, *SD*=3.2, *Mdn*=24); this difference also showed large practical significance (*U*=4.0, *Z*=-2.84, *p*=.003, $r_{rb}$=.87).

The ability to use referential language was measured at 18 and 24 months with the MB-CDI use of words measure. At the age of 18 months, the mean score of the use of words in reports from twins' parents was 0.9 lower (*M*=3.5, *Mdn*=2.5, *SD*=2.4) than the reported mean in the method manual (*M*=4.4, *Mdn*=4.0, *SD*=2.4). No significant group differences were found in any of the background variables for word use at 18 months (see all statistical information in Appendix 19). The difference of means between twin data and normative data also remained small in the use of words at 24 months. Twins' mean score was 0.8 lower (*M*=7.4, *SD*=1.9, *Mdn*=8.0), when compared with information in the method manual (*M*=8.2, *SD*=2.1, *Mdn*=5.0).

At 24 months, two statistically significant group differences were discovered for the scores in MB-CDI referential use of words (see Appendix 20 for all statistical information). The use of words score was found to be lower for twins born with birth complications (*M*=5.7, *Mdn*=5.5, *SD*=1.2) than for twins born without birth complications (*M*=8.3, *Mdn*=9.0, *SD*=1.5). This difference showed large practical significance (*U*=6.5, *Z*=-2.80, *p*=.003, $r_{rb}$=.82). In addition, the score of the use of words remained lower for twins with lower-educated fathers (*M*=6.3, *SD*=1.7, *Mdn*=6.0), when compared with twins with higher-educated fathers (*M*=8.88, *SD*=0.83, *Mdn*=9.0) – a difference showing large practical significance (*U*=9.0, *Z*=-2.8, *p*=.004, $r_{rb}$=.78). No other group differences were detected in any demographic background variables for referential use of words at 24 months.

The development of twins' syntactic and grammatical skills was measured with the MB-CDI WS sum variable of inflectional morphology and MSL at 18 and 24 months. Figure 12 presents the distributions of twins' norm-adjusted scores of morphological complexity at the ages of 18 and 24 months. Twins seemed to be less mature in language skills related to morphological complexity, when compared with norm data (*M*=2.7, *Mdn*=2.0, *SD*= 2.9). At 18 months, the mean complexity score of twins was 0.7, with a median of 0.5 and the standard deviation of 1.0 (range=0–3). Preterm twins (*M*=1.5, *Mdn*=1.3, *SD*=1.1) in this study were more skilled than term children (*M*=0.2, *Mdn*=0, *SD*=0.3); this difference showed substantial practical significance (*U*=11.0, *Z*=-2.99, *p*=.003, $r_{rb}$ =.77). No other statistically significant group differences between any demographic background variables in morphological complexity were found at 18 months (see also Appendix 19).

Twins' proficiency in morphological complexity also showed differences in regards to normative data at 24 months of age. Twins' mean score was 6.5 (*Mdn*=6.8*,* *SD*=3.5), which falls 2.8 points below the MB-CDI normative mean of morphological complexity of 9.3 (*Mdn*=10.0, *SD*=5.0). The statistically significant group difference was evident between preterm and full term children also at 24 months. Preterm children continued to have more morphological proficiency (*M*=8.9, *Mdn*=9.0, *SD*=0.5), when compared with full-term children (*M*=5.3, *Mdn*=5.0, *SD*=3.7); this difference showed substantial practical significance (*U*=12.0, *Z*=-226, *p*=.024, $r_{rb}$=.67). In addition, twins with older siblings were more advanced in morphological skills (*M*=8.3, *Mdn*=9.0, *SD*=2.8), when compared with first-born twins (*M*=4.4, *Mdn*=4.0, *SD*=3.2) – a difference also having substantial practical significance (*U*=14.5, *Z*=-2.28, *p*=.021, $r_{rb}$=.64). No statistically significant differences were found for the sum variable of morphology in other neonatal health-related variables or any of the demographic background variables (Appendix 20).

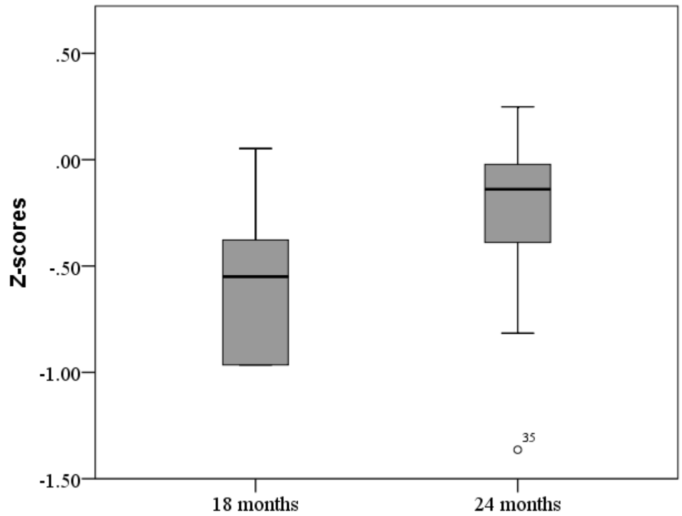



**Figure 12.** Morphology scores of twins adjusted to Finnish MB-CDI norms

Twins' maximum sentence length (MSL) was on average slightly shorter than those of children in the norm study at both ages (Figure 13). The difference did, however, increase with age. In the norm sample, mean MSL scores of 18-month-olds were 1.8 (*Mdn*=1.0, *SD*=1.4) in normative data, while the mean score for 18-month-old twins was 1.4 (*Mdn*=1.2, *SD*=0.6) in this data. At 24 months, the normative mean score was 5.7 (*Mdn*=5.7, *SD*=3.0), while in this data, twins' mean MSL at 24 months was 4.6 (*Mdn*=4.5, *SD*=2.1). No significant differences were found in group comparisons between any of the background variables at 18 months (see Appendix 19). However, children with older siblings were found to produce more complex utterances at 24 months (*M*=5.5, *Mdn*=5.3, *SD*=1.7), when compared with utterances from first-born twins (*M*=2.9, *Mdn*=3.3, *SD*=1.6); this difference showed a large practical difference (*U*=4.0, *Z*=-2.84,

$p$=.003, $r_{rb}$ =.87). No other statistically significant group differences were found between twins' MSL scores at 24 months (Appendix 20).

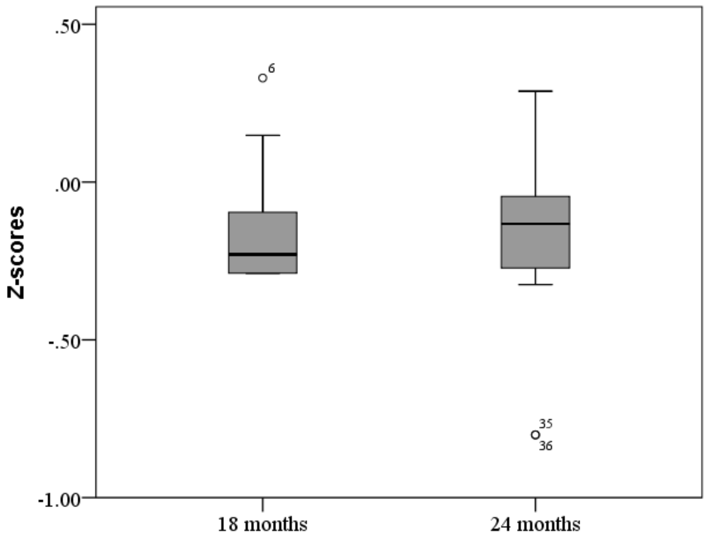



**Figure 13.** MSL of twins adjusted to Finnish MB-CDI norms

### 5.3.4 Summary of the vocal milestones, vocabulary, and linguistic skills of twins

Twins started reduplicative babbling within normal variation, but the onset of variegated babbling differed substantially from reported normative information. The onset of variegated babbling was delayed significantly, when compared with singleton normative data. A-twins started reduplicated babbling earlier than B-twins, and protowords were discovered earlier in LBW twins and in twins with higher-educated mothers, but no other statistically significant group differences were detected between background variables and milestones.

Twins scored lower in vocabulary at 12, 18, and 24 months, when compared with the normative sample, but their scores remained within normal variation. No constant statistically significant group differences were evident for any of the background variables from 12 to 24 months. However, there were single statistically significant differences for several background variables at 12 months: 1) children from families with the father having a lower education had smaller receptive and expressive vocabularies,

2) children born with complications understood fewer words than children born without complications, and 3) children of low birth weight produced more words than children with normal birth weight. In addition, preterm twins produced slightly more words at 18 months, when compared with full-term twins, and twins with siblings had substantially larger expressive vocabularies than first-born twins at 24 months.

On average, the early signs of understanding and the use of communicative actions and gestures showed to be similar in twins as in the normative data. In the group comparisons, children born with birth complications had lower scores in the first signs of understanding, and first-born twins had lower scores in communicative actions and gestures, when compared with children with older siblings. Additionally, the sum of actions, gestures, and non-verbal communication abilities showed to be more advanced in children from families with the mother having a lower education.

At 18 and 24 months, all the language scores (use of words, sum of morphology, MSL) were lower for twins, when compared with the norm data, but remained within normal variation. No statistically significant group differences were found for the use of words and MSL at 18 months, but preterm twins were found to be advanced in morphological skills, when compared with full-term twins. Other group differences evident at 24 months were: 1) the lower use of words score for children born with complications and from lower-educated fathers, 2) better morphological scores of twins with siblings, when compared with first-born twins, and 3) a better MSL score in twins with siblings, when compared with first-born twins.

## 5.4 Uniformity of measures: is there a relationship between child volubility, quantitative language environment, and the acquisition of early language skills?

In the study, concerning questions related to the uniformity of the automated measures, the interest was both in the associations between the volubility of different speakers and between LENA-provided variables and parent-reported variables. In the following chapter, the intercorrelations between LENA speaker segment durations are presented, followed by the results related to the associations between speaker segment durations and parent-reported onset of vocal milestones, vocabulary, and language skills. Chapter 5.4.2 will focus on the internal correlations within CDEV and MB-CDI and present results from the inspections of the relationship between parent-reported onset of vocal milestones, vocabulary, and language skills.

### 5.4.1 The relationship of the LENA speaker segment durations and parent-reported pre-lexical milestones, vocabulary, and language skills from CDEV and MB-CDI

Correlational analysis showed that there was a positive statistically significant association between the median durations of female and male adult speaker categories (Table 24). but no statistically significant associations were found between key child and other child, in relation to any of the speaker categories. Although female and male adult volubility was found to be correlated to each other, there was no association between any of the speaker segment durations and parent-detected onset of reduplicated babbling. These findings suggest that the talkativeness of one parent is associated with the talkativeness of the other parent, but not with the talkativeness or volubility of children in the family. However, in this data, the mean volubility of male adults was found to be strongly associated with the detected onset of variegated babbling, and the mean volubility of female adults was moderately associated with the detection of protowords. The more male adults were talking, the later the twins were discovered to start varying babbled utterances. In addition, the more the female adults were talking, the later the twins were found to produce first protowords.

**Table 24.** Correlations between LENA speaker segment durations, CDEV milestones, and MB-CDI vocabulary scores

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Female adult mean durations | | | | | | | | | | |
| 2. Male adult mean durations | **.59**** | | | | | | | | | |
| 3. Key child mean durations | -.39 | -.11 | | | | | | | | |
| 4. Other child mean durations | -.17 | .36 | -.11 | | | | | | | |
| 5. Onset of reduplicated babbling | .11 | .21 | .06 | -.07 | | | | | | |
| 6. Onset of variegated babbling | .04 | **.84**** | .48 | .22 | .24 | | | | | |
| 7. Emergence of protowords | **.58*** | .04 | -.17 | -.42 | .06 | -.37 | | | | |
| 8. 12-month receptive vocabulary | .29 | .05 | **-.50*** | .00 | -.01 | -.08 | .32 | | | |
| 9. 12-month productive vocabulary | -.41 | -.31 | .24 | -.09 | .30 | -.14 | -.24 | .19 | | |
| 10. 18-month productive vocabulary | -.05 | -.06 | -.03 | .19 | -.16 | -.43 | .35 | **.59**** | **.45*** | |
| 11. 24-month productive vocabulary | .32 | **.51*** | -.09 | .32 | -.03 | .24 | .37 | **.50*** | .15 | **.71**** |

p<.05*, p<..01**

In addition, the larger amount of key child vocal activity was associated with a lower amount of words understood at 12 months. However, although most relations suggested a negative association between the amount of family member talk and twin development, a moderate positive correlation was found between the mean amount of male adult talk and the amount of productive vocabulary at the age of 24 months.

When overall mean speaker segment durations were analyzed with MB-CDI language scores using Spearman's non-parametric correlation analysis, no statistically significant associations were found between female adult volubility and MB-CDI language scores (Table 25). However, female adult and male adult mean durations were positively and statistically significantly related to more advanced non-verbal communicative gestures. Other child overall segment durations were positively and statistically significantly associated with twins' maximum sentence length at 24 months of age. However, no statistically significant associations were detected between any of the speaker segment durations and MSL at 18 months, nor between the referential use of language or with morphological skills at 18 and 24 months.

**Table 25.** Correlations between LENA segment durations and MB-CDI language scores

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Female adult mean durations | | | | | | | | | | | |
| 2. Male adult mean durations | **.59**** | | | | | | | | | | |
| 3. Key child mean durations | -.39 | -.11 | | | | | | | | | |
| 4. Other child mean durations | -.17 | .36 | -.11 | | | | | | | | |
| 5. First signs of understanding at 12 months | .02 | .11 | **-.46*** | .02 | | | | | | | |
| 6. Actions and gestures at 12 months | **.55*** | **.70**** | -.04 | .43 | -.12 | | | | | | |
| 7. Use of words at 18 months | .08 | -.07 | -.20 | .01 | .28 | .32 | | | | | |
| 8. Morphology at 18 months | .12 | .23 | .16 | -.08 | .08 | **.59*** | **.63**** | | | | |
| 9. MSL at 18 months | -.21 | -.21 | -.12 | .19 | .19 | .19 | .36 | .20 | | | |
| 10. Use of words at 24 months | -.07 | .04 | .12 | 0 | .38 | -.22 | .14 | .06 | .26 | | |
| 11. Morphology at 24 months | -.15 | .44 | .13 | .41 | -.13 | **.68**** | .32 | **.66**** | .29 | .07 | |
| 12. MSL at 24 months | -.21 | .20 | -.13 | **.59*** | .14 | **.65*** | **.85**** | **.63**** | .41 | .10 | **.75**** |

p<.05*, p<.01**

### 5.4.2 The relationship of LENA core measures and parent-reported pre-lexical milestones, vocabulary, and language skills in CDEV and MB-CDI

When the LENA core measures from the pooled data were analyzed with Spearman's non-parametric correlational analyses, statistically significant moderate correlations were found between mean AWC and CTC counts and mean CVC and CTC counts, but no statistically significant correlations were found between mean AWC and CVC counts (Table 26). A statistically significant moderate negative association was detected between AWC and 12-month productive vocabularies, suggesting that the more adult words LENA has detected, the fewer words the children have produced at 12 months. In addition, strong positive associations between CTC and the onset of protowords suggest that when the amount of LENA CTC is large, the onset of protowords is detected later. No other statistically significant correlations were detected for LENA AWC and CTC, and no correlations were detected between LENA CVC and milestones and vocabularies measured with CDEV and MB-CDI.

**Table 26.** Correlations between LENA speaker segment durations, CDEV milestones, and MB-CDI vocabulary scores

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1. AWC | | | | | | | | | |
| 2. CVC | -.12 | | | | | | | | |
| 3. CTC | **.63**** | **.46*** | | | | | | | |
| 4. The onset of reduplicated babbles | .08 | -.-40 | -.16 | | | | | | |
| 5. The onset of variegated babbles | .43 | 0.30 | -.19 | .20 | | | | | |
| 6. The emergence of protowords | .47 | .20 | **.80**** | .10 | -.40 | | | | |
| 7. 12-month receptive vocabulary | .29 | -.20 | .26 | 0 | -.10 | .30 | | | |
| 8. 12-month productive vocabulary | **-.47*** | -.20 | -.44 | .30 | -.10 | -.20 | .19 | | |
| 9. 18-month productive vocabulary | -.07 | .01 | .13 | -.20 | -.40 | .40 | **.59**** | **.45*** | |
| 10. 24-month productive vocabulary | .44 | -.20 | .22 | 0 | .20 | .40 | **.50*** | .15 | **.71**** |

p<.05*, p<.01**

When overall LENA core measure counts were analyzed with MB-CDI language scores using Spearman's non-parametric correlation analysis, the only statistically significant correlation between LENA core measures and MB-CDI language scores was found for LENA AWC and the 12-month sum score of communicative actions and gestures (Table 27). The moderate positive correlation suggested that when the amount of mean adult words increased in the pooled data, the children's use of communicative actions and gestures was also more versatile. No other statistically significant correlations were found for AWC, and no statistically significant correlations were found between any of the CVC and CTC counts and MB-CDI language scores.

**Table 27.** Correlations between LENA core measures and MB-CDI language scores

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| 1. AWC | | | | | | | | | |
| 2. CVC | -.12 | | | | | | | | |
| 3. CTC | **.63**$^{**}$ | **.46**$^{*}$ | | | | | | | |
| 4. Actions and gestures 12 months | **.60**$^{*}$ | -.20 | .23 | | | | | | |
| 5. Use of words at 18 months | .01 | -.09 | .15 | .32 | | | | | |
| 6. Morphology at 18 months | .19 | .04 | .22 | **.59**$^{*}$ | **.63**$^{**}$ | | | | |
| 7. MSL at 18 months | -.15 | .00 | -.19 | .19 | .36 | .20 | | | |
| 8. Use of words at 24 months | .06 | .23 | .13 | -.22 | .14 | .06 | .26 | | |
| 9. Morphology at 24 months | .06 | -.18 | -.31 | **.68**$^{**}$ | .32 | **.66**$^{**}$ | .29 | .07 | |
| 10. MSL at 24 months | -.04 | -.25 | -.18 | **.65**$^{*}$ | **.85**$^{**}$ | **.63**$^{**}$ | .41 | .10 | **.75**$^{**}$ |

$p<.05^{*}$, $p<.01^{**}$

### 5.4.3 The relationship of parent-reported milestones in CDEV and early vocabulary and language skills in MB-CDI

The associations between parent-reported vocal milestones and later vocabulary was studied using non-parametric Spearman correlations (Table 28). No statistically significant associations were detected between parent-reported onset of reduplicated babbling and vocabulary, parent-reported onset of variegated babbling and vocabulary, nor parent-reported discovery of protowords and vocabulary at 12, 18, or 24 months. In addition, in this data, the parent-reported vocal milestones collected with CDEV were not correlated with each other, suggesting that there was no statistically significant associations between the onset of reduplicated babbling, the onset of variegated babbling, and the detected emergence of protowords. However, MB-CDI vocabularies at the ages of 12, 18, and 24 months showed constant moderate to strong internal correlations, with the exception of 12-month productive vocabulary, which was only correlated to vocabulary at 18 months.

When parent-reported milestones from CDEV were studied in relation to MB-CDI language scores, no statistically significant associations were found between milestones and any of the language measures from MB-CDI (Table 29). In addition, the MB-CDI score from the first signs of understanding at 12 months, MSL at 18 months, and the referential use of words at 24 months did not show statistically significant associations with any of the CDEV or MB-CDI language scores. There were, however, several internal correlations within the MB-CDI language scores. At 12 months, the sum of actions and gestures was associated with parent-reported language skills on morphology at 18 and 24 months and with MSL at 24 months as well. Additionally, the referential use of words and MSL at 18 months were strongly associated with each other, but with MSL at 24 months as well. In addition, the sum of morphology and MSL at 24 months were strongly correlated with each other.

**Table 28.** Correlations between prelexical milestones and vocabulary at 12, 18, and 24 months

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. the onset of reduplicated babbling | | | | | | |
| 2. the onset of variegated babbling | .24 | | | | | |
| 3. the onset of protowords | .06 | -.37 | | | | |
| 4. 12-month receptive vocabulary | -.01 | -.08 | .32 | | | |
| 5. 12-month productive vocabulary | .30 | -.14 | -.24 | .19 | | |
| 6. 18-month productive vocabulary | -.16 | -.43 | .35 | **.59**$^{**}$ | **.45**$^{*}$ | |
| 7. 24-month productive vocabulary | -.03 | .24 | .37 | **.50**$^{*}$ | .15 | **.71**$^{**}$ |

p<.05*, p<.01**

**Table 29.** Correlations between prelexical milestones and MB-CDI language scores

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Reduplicated babbling | | | | | | | | | | |
| 2 | Variegated babbling | .24 | | | | | | | | | |
| 3 | Protowords | .06 | -.37 | | | | | | | | |
| 4 | First signs of understanding at 12 months | .03 | -.31 | -.25 | | | | | | | |
| 5 | Actions and gestures 12 months | -.23 | .57 | .23 | -.12 | | | | | | |
| 6 | Use of words at 18 months | -.09 | -.41 | .40 | .28 | .32 | | | | | |
| 7 | Morphology at 18 months | -.04 | .19 | .37 | .08 | **.59**$^{*}$ | **.63**$^{**}$ | | | | |
| 8 | MSL at 18 months | -.27 | -.24 | -.22 | .19 | .19 | .36 | .20 | | | |
| 9 | Use of words at 24 months | .01 | .51 | -.26 | .38 | -.22 | .14 | .06 | .26 | | |
| 10 | Morphology at 24 months | -.16 | .47 | -.16 | -.13 | **.68**$^{**}$ | .32 | **.66**$^{**}$ | .29 | .07 | |
| 11 | MSL at 24 months | -.30 | -.17 | .11 | .14 | **.65**$^{*}$ | **.85**$^{**}$ | **.63**$^{**}$ | .41 | .10 | **.75**$^{**}$ |

p<.05*, p<.01**

Internal correlations between MB-CDI scores to vocabulary and language show moderate to strong statistically significant associations between many of the measures (Table 30). The most constant variables with associations to other variables were the 12-month receptive vocabulary, the 18-month productive vocabulary, and the 18-month sum of morphology. The fewest associations to the 24-month outcome scores were found in the 12-month expressive vocabulary and the 18-month MSL. In addition, the variable with the most associations from 24 months to preceding measures was MSL at 24 months.

**Table 30.** Internal correlations between MB-CDI vocabulary and language scores

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. 12-month receptive vocabulary | | | | | | | | | | | |
| 2. 12-month productive vocabulary | .19 | | | | | | | | | | |
| 3. Vocabulary at 18 months | **.59**$^{**}$ | **.45**$^{*}$ | | | | | | | | | |
| 4. Vocabulary at 24 months | **.50**$^{*}$ | .16 | **.71**$^{**}$ | | | | | | | | |
| 5. First signs of understanding at 12 months | **.52**$^{*}$ | .01 | .20 | 0 | | | | | | | |
| 6. Actions and gestures at 12 months | .06 | -.33 | .35 | **.74**$^{**}$ | -.12 | | | | | | |
| 7. Use of words at 18 months | **.67**$^{**}$ | .42 | **.91**$^{**}$ | **.63**$^{**}$ | .28 | .32 | | | | | |
| 8. Sum of morphology at 18 months | .30 | .14 | **.58**$^{**}$ | **.62**$^{**}$ | .08 | **.59**$^{*}$ | **.63**$^{**}$ | | | | |
| 9. MSL at 18 months | .34 | .12 | .36 | .3 | .19 | .19 | .36 | .2 | | | |
| 10. Use of words at 24 months | **.54**$^{*}$ | .21 | .18 | .15 | .38 | -.22 | .14 | .06 | .26 | | |
| 11. Sum of morphology at 24 months | .15 | .22 | .39 | **.66**$^{**}$ | -.13 | **.68**$^{**}$ | .32 | **.66**$^{**}$ | .29 | .07 | |
| 12. MSL at 24 months | **.56**$^{*}$ | .24 | **.94**$^{**}$ | **.73**$^{**}$ | .14 | **.65**$^{*}$ | **.85**$^{**}$ | **.63**$^{**}$ | .41 | .10 | **.75**$^{**}$ |

$p<.05^{*}$, $p<.01^{**}$

### 5.4.4 Summary on the uniformity of information from automated measures and parent reports

When the association of LENA speaker segments and CDEV milestones was studied, it was found, that the more male adults were found to talk, the later the twins were discovered to start varying babbled utterances. In addition, the more the female adults were talking, the later the twins were found to produce first protowords.

The increased amount of key child volubility was associated with MB-CDI lower amount of words understood at 12 months. However, the amount of male adult and key child volubility was positively associated with the children's non-verbal communicative skills at 12 months. In addition, a moderate positive association was found between the mean amount of male adult talk and the amount of productive vocabulary at the age of 24 months.

The increased amount of conversational turns in this study was associated with later emergence (or detection) of protowords. The increased amount of adult words was positively associated with children's abilities to use non-verbal communicative actions and gestures at 12 months, but the increased amount of adult words showed negative associations with a 12-month productive vocabulary. No other statistically significant associations were found between any of the LENA measures and CDEV or MB-CDI measures.

The study of internal correlations within LENA speaker segments found that the amount of male adult and female adult segments were associated with each other, but not with the amount of key child and other child speech. In addition, the amount of adult words and the count of child vocalizations were associated with the amount conversational turns, but not with each other. No internal associations were detected between CDEV vocal milestones, but several associations were found within MB-CDI vocabulary and language scores at the ages of 12, 18, and 24 months.

# 6 Discussion

The main findings of this study include results on twins' language development, results related to neonatal and social environment of twins, and the reliability of the studied LENA method. For twins' language acquisition, the present study suggests that although twins scored lower on vocabulary, referential language, and morphosyntactic skills, when compared with normative information, their performance remained within normal variation. However, the results show variation within the studied subgroups of twins in relation to neonatal health and social environmental factors.

The amount of family speech showed some positive and some negative associations to twins' development. The increased amount of adult talk was negatively associated with the parents' detection of variegated babbling and protowords, but the amount of adult speech was positively associated with children's non-verbal communication skills. Interestingly, the amount of heard male adult speech was positively associated with twins' mastery of vocabulary at the age of 24 months, and the amount of sibling speech was positively associated with maximum sentence length at the age of 24 months. In addition, the group comparisons of background variables showed that twin children from families with older siblings heard more male adult and sibling speech, had larger vocabularies at the age of 24 months, and more advanced morphosyntactic skills, when compared with twins without siblings.

In addition to child development, the study showed that the identification of female adults and children was sufficient in LENA, and the key child vocalization count of the system was reliable. However, the detection of male adult identification and the calculations of adult word counts did not show to be reliable. In addition, it is to be noted that although group comparisons showed the gender and the level of neonatal care to produce statistically significant group differences in the amount of family member's talk, neither gender, nor the level of neonatal care showed to have importance in twins' language development using the current measurements.

## 6.1 Examination of the results

The current study was two-fold in nature, focusing on questions related to the methods and questions related to twins' language acquisition. A novel method was assessed, used, and information was derived from it; the method was compared with information gained through more traditional methods used in clinical practices. In addition, the data from twin children's language development was compared with the available normative data in order to provide additional information on twins to build on the existing knowledge of the development of children's language in general. Therefore, the discussion will start with considerations related to the results on LENA reliability, followed by the discussion of results about the early environment and language acquisition of twins. Lastly, the discussion of results will be focused on the associations and a lack of associations between the measures and the measured variables in this study.

### 6.1.1 The implications of suitability of LENA in Finnish

This study is one of the first studies to explore LENA with non-English data and the first to discuss its suitability and use in Finnish. The work presented findings from analyses of two all-day recordings (R1 and R2) and acknowledges that future work with a more comprehensive sample is needed, if LENA was to be validated for Finnish.

Although the sample for the reliability trial only had two recordings, the data had over 100 000 segments. By listening, checking the segment, labelling, and transcribing all of the audio, it was possible to learn to understand the data well. Therefore, it became possible to put some of the aspects, which may have influenced the results, up for discussion. It is also to be noted that both of the recordings were conducted in twin families, which set the program and the human gold-standard, facing the particularly hard task of being required to distinguish between two same-aged children living in the same household. Despite the twin situation, the results for segmentation and speaker identification were mainly encouraging, although the identification of male adult segments did not reach a satisfactory level in R1. This is an aspect that will require further studying on its own, but also, further investigations of the LENA conversational turns count measure, as the conversational turn count relies on accurate key child, female, and male adult segmentation (Xu, Yapanel, Gray, Gilkerson, Richards & Hansen, 2008).

The inter-rater procedure yielded high rates of agreement for both key children, differing substantially from the agreement rates from LENA technical reports. Agreement rates were found to be good to very good for female adult and other child segments as well, whilst the lowest agreement was found for male adults. The lower segment accuracy for males is not in accordance with previous studies from VanDam and Silbert (2013a), who found LENA to distinguish male adults better than female adult segments. However, in this study, the male adult segment sensitivity, specificity, and accuracy were found to be high in both of the recordings. This led to the question of why the overall accuracy was distinctively high for adult males, if the agreement rates turned out to be low.

It was found that LENA does a fairly good job of identifying the true male negatives, while producing only a small number of false negatives. Thus, we can reason that in this sample, LENA errors for males were mainly false positives; this view is also supported by the high false discovery rate. With the experience from transcription, it is suggested that the majority of false positives for male adults in this data might arise from two origins: 1) the mix-ups in automatic labelling of female-produced lower-pitch utterances and 2) the inability of the algorithm to distinguish differences between present male and electronic male sounds coming from a radio – a question also raised for discussion by Blackwell, Babayigit, and Roulstone (2013). Although both suggestions would require further investigations, the author wants to note a few of the aspects, which might explain the large number of false positives for male adults.

Firstly, the lowest sensitivity rate for speakers was found to be the rate of adult females in the first recording (R1). This is largely explained by the same misclassifications explaining the false positives for male adults. The misinterpretations, lowering sensitivity for female adult in R1, were in fact false negatives, accumulating in the male adult class. This does not, however, affect the false discovery rate, nor total accuracy, which is very good for female adults both in R1 and in the second recording (R2). Secondly, there are suggestions, stating that fundamental frequencies (F0) are culturally related, and that Finnish females would have lower F0, when compared with English-speakers (Guimarães & Gouveia, 2007). On the other hand, CDS is characterized by high-frequency speech, and it has been reported to bear cross-linguistic similarities (Fernald, Taeschner, Dunn, Papousek, Boysson-Bardies & Fukui, 1989; Trainor & Desjardins, 2002). Thirdly, although LENA is constructed to detect electronic sounds, the reliability of the detection should be studied further, as it was recognized as one important source of error in the current study.

Because there is very little research on the reliability of detection of electronic sounds in LENA, this needs to be taken into consideration, when planning future studies. Some pilot work in measuring TV exposure and its effect on child development has already been published (Christakis *et al.*, 2009), and the measure could be much appreciated in clinical use. On the other hand, if the identification of electronic sounds does not reach satisfactory levels of reliability, a reduction of electronic sound sources from recording environments could enhance the reliability of male speaker detection. However, this would lessen the naturalistic nature of the recordings, as TV, radio, and other electronic devices in many families are a part of their everyday life.

As stated earlier, the agreement rates for key child segments showed very good agreement between the machine and human coders, as was sensitivity in both of the recordings. Importantly, LENA was also precise in key child detection, which does differ from the study conducted with the Chinese Mandarin Dialect. Gilkerson *et al.* (2014) also found key child sensitivity to be good (81%) in LENA, but the precision of the system remained low (27%). However, some aspects in key child detection need further discussion. One finding was that specificity for key children turned out to be lower, especially in analyses from R1, which was recorded in a family with an older sibling. It is suggested that the lower rate of specificity is mainly explained with false positives for the key child in R1, but the number of false positives does not affect total accuracy because of the high number of true positives and true negatives and the low number of false negatives. This finding is also supported by a fairly low false discovery rate, suggesting that LENA can judge key child labels from R1 in an adequate manner.

Another point of discussion is the overall accuracy and its relation to false discovery rate in R2 key child and other child labels. Overall accuracy for key child in R2 yielded a fairly good rate of 84%, but the false discovery rate rose to 18%, suggesting that LENA made some errors to account for the false positives for R2 key child labelling. This might be explained by misinterpretations of key child segments in R2 other child segments, which produced most of the false positives for R2 key child. When focusing on R2 other child labels, LENA was found to be accurate and specific, but sensitive to a lesser extent. Lower sensitivity is likely to have been affected by the substantial number of false negatives labelled as R2 key child instead R2 other child. However, the misinterpretation between key and other child does not fully explain the false positives affecting the fairly high false discovery rate of 30% for R2 other child. Rather, false discoveries are accumulated, when LENA was misinterpreting other child labels with a non-existent speaker (not categorized) or with the key child and female adult categories.

LENA labels key child segments according to four categories: speech-like vocalizations, vegetative sound, cry, and silence. Segments are labelled "silence", if a pause of 300 ms is present in child-produced utterances (Oller *et al.*, 2010, supporting material). Therefore, it is natural that silent segments were nearly perfectly detected by LENA, as shown in the high level of agreement rate, sensitivity, specificity, accuracy, and low false discovery rates. However, the remaining three categories need closer examination.

Vegetative sounds (burps, hic-ups, laughter, etc.) are thought to be easily distinguishable from vocal sounds, such as speech-like vocalizations and cries, because of the clear differences in their acoustic properties (for LENA identification of speech-like vocalizations, see the supporting information in Oller *et al.*, 2010). In this study, percent agreement rates, specificity and accuracy towards vegetative sounds were found to be high. However, LENA was not highly sensitive to vegetative sounds. The reason for this lies in the number of false negatives, which were categorized as cry sounds by the program, but detected as vegetative sounds by the human. But, as false negatives do not account for false discovery rate, FDR was found to also be fairly good for vegetative sounds.

Unlike silence and vegetative sounds, inter-rater agreements turned out lower for cries in both of the recordings, with a substantial number of false positives that were labelled as speech-like vocalizations by the primary human coder. In the author's experience, this interpretational difference most often occurred with vocalizations in particular, which were not clear babblings, but were produced with a passive articulatory tract and/or had a slight implication of unhappiness, instead of clear dissatisfaction or cry. This interpretational difference had an effect on false discovery rates for cry sounds, as well as on the number of false negatives for speech-like vocalizations.

Although the previously mentioned false negatives affected sensitivity and accuracy, in this study, it was found that LENA identified speech-like vocalizations well. The agreement rates of machine and human were good, as was specificity and total accuracies. LENA was less sensitive in detecting speech-like vocalizations, but false discovery rates were adequately low, meaning that LENA did not make false positive speech-like vocalization errors within key child segments. This is particularly important because some LENA features (e.g. CTC; Automatic Vocalization Assessment, AVA™) partially rely on segmentation and labelling (Xu, Yapanel, Gray, Gilkerson, Richards, Hansen, 2008).

The algorithms of the LENA System have been trained with American-English data, and therefore, are to some extent language-specific. However, the three previous studies, where AWC was evaluated for non-English (Chinese SDM, French, and Spanish), concluded that LENA was able to estimate adult word counts in an adequate manner (Gilkerson *et al.*, 2014; Canault *et al.*, 2015; Weisleder & Fernald, 2013). This claim was not confirmed for Finnish in the current research, possibly because of a different way of studying the reliability. Previous studies have used correlations to study the relation of human and machine counts, but for this study, the difference between machine and human was analyzed with absolute and relative error rates. This selection was based on the notion that correlational analysis may not be the best way of assessing agreement, and thus, the interpretation of the results would be vulnerable to errors (Bland & Altman, 1986). Although the decision to look for absolute and relative differences showed to be beneficial towards understanding the unsystematic nature of estimation error in the current study, future studies could perhaps apply the 95% confidence intervals suggested by Bland and Altman (1999).

As with the segment reliability trials, the sample in the reliability inspection of LENA adult word count (AWC) and child vocalization count (CVC) was limited to only two families. However, the results suggest that further exploration is needed before relying on LENA adult word count reports on Finnish data, especially since the substantial difference between LENA and human word counts could not be resolved with an exploration of relative error rates. Error rates were not consistent enough to be resolved as simply as applying a certain coefficient to correct the discrepancy in AWC counts. In the current study, error rates for AWC varied greatly, but in LENA technical reports, AWC has been reported to be on average 2% lower than the mean count from human transcribers (see further Xu *et al.*, 2008; LTR-05-02).

The reasons for differing word counts may partially lie in false positives, where LENA mistakenly identifies speech from TV or radio as a male or female speaker. However, additional explanations may well lie in the contrastive differences between American-English and Finnish. Although the basis of LENA core counts lies in speaker segment identification, they are further analyzed with statistical models, which have been taught with transcription data from American-English child and adult samples. Previously, Gilkerson *et al.* (2014) stated that AWC is the most dependent part of American English-based modeling within LENA measures, as it relies on segment duration, silence distribution, and phone-based consonant and vowel distributions, which differ between the two languages. For example, monosyllabic words are frequent in AE, whereas they are less common in Finnish (Saaristo-Helin, Kunnari & Savinainen-Makkonen, 2011). On the other hand, Finnish is an agglutinative language and may thus have longer word structures due to complex morphology, when compared with English (Vannest, Bertram, Järvikivi & Niemi, 2002). In Finnish, additional meanings are attached to words most often with suffixes (e.g. <talo+i+ssa+mme+ko>; "In our houses?"), and compound words are also common (e.g. <villa + haalari> "An overall made of wool") (for more detailed presentation of Finnish, see e.g. Helasvuo, 2008 Saaristo-Helin, Kunnari & Savinainen-Makkonen, 2011).

The explanation that linguistic differences could account for errors in adult word counts may be supported by the only slight and more consistent underestimation of child vocalization counts by LENA. Prelexical child vocalizations are much more universal in nature, and hence, their counts should be less language-dependent. However, another explanation for minor and consistent errors might lie in the fairly good performance of LENA key child segment identification, as the CVC is built upon segment identification. The encouraging finding of LENA CVC reliability is also in accordance with the study of Soderstrom and Wittebolle (2013), who also reported a slight underestimation of CVC counts from family home data. However, as in the current study, the key children were babblers; it is recommended that in future studies, the reliability of child vocalization counts should be explored with more language-specific data gathered from older, already talkative children.

### 6.1.2 Language environment in twin families

This study used LENA to quantify and describe the amount of speech heard by twins in their everyday family environment, which has been suggested to be of importance in regards to twin language development in previous literature (Lytton *et al.*, 1977; Tremblay-Leveau *et al.*, 1999). Two types of variables were measured: 1) the overall time female adults, male adults, and other children vocalized or spoke in the presence of twins according to LENA, and 2) the core measures of adult words (AWC) spoken near the child and conversational turns involving the key child (CTC). The decision to take segment durations as variables in addition to AWC and CTC was based on reliability checks, as LENA segment identification showed better performance than AWC counts, which partially affect CTC counts as well.

As this study did not include any explicit reliability checks on CTC, interpretations of CTC needs to be cautious and conservative. Although it could have been justified to leave the AWC and CTC information out of the present study, the author made the decision to report AWC and CTC results in order to participate in generating the close to non-existent body of information gathered with the LENA tool. This decision was based on the fact that more information about the applicability of core measures is needed, as LENA is already being applied in pilot-studies in non-English language areas (e.g. Gilkerson *et al.*, 2014; Löfkvist, 2016).

Mean segment durations in 12-hour adjusted counts showed that the proportion of clear "meaningful" speech and vocalizations was less than 20% of the mean total duration of the recordings. This is, of course, not all of the produced speech in close proximity to twins, nor does it represent the amount of child-directed speech, but, if further studied, might serve as an implication of the proportion of clear speech surrounding the children. As LENA cannot identify meaningful speech from segments with overlapping sounds, the fairly low proportion of meaningful segments may be partially explained with the locomotive developmental phase of the studied children. During the study period, children became active crawlers, who were only in the last few recordings learning to manage their upright position. Therefore, the majority of recordings contained an unclear signal due to the way the child was able to move. In the latter recordings, however, the source for overlapping signals might be explained (at least partially) by the motor development of the child, although the upright position was already established. In the later recordings, children were in the developmental stage of having repetitive hand movements and were, thus, banging toys, waving their arms and creating a lot of fuzz and noise in the recordings.

In the present work, AWC counts suggested that Finnish-acquiring twins heard spoken words somewhat similarly to children in the LENA normative data, although AWC had a rising trend in twin family data. This trend was also present for CTC counts, but with a modest correlation between conversational turn counts and the age of the child. These results are not in accordance with previous results, suggesting that parents of twins would talk less to their children, when compared with singletons (Conway *et al.*, 1980; Holditch-Davis *et al.*, 1999). However, although the amount of all speech could be similar in close proximity to twins and singletons, it is likely that the amount of child-directed speech to twins may be less than the amount of child-directed speech to singletons, as suggested by Tomasello and colleagues (1986). In addition, it needs to be emphasized that the current results are of relevance only if there is reason to assume that the algorithm in LENA works properly enough with Finnish data. Unfortunately, the reliability trial does not support this assumption for AWC and, as CTC relies partly on AWC, the reliability of the CTC may also be questionable. On the other hand, monthly measurements of volubility showed a significant increase in female adult speech segments and also a moderate increase in male adult segments. This information supports the true increase of AWC and CTC. An additional angle to the matter is provided in chapter 6.1.4, where the results of correlational analyses between LENA measures and parent reports are discussed.

This study had four neonatal health-related variables, which were used to inspect the possible effects of children's early medical problems and the effect of neonatal care on the interaction occurring in families. All the variables were selected based on the literature review (see 2.1.2) and the widely reported effects of pretermity and medical risks in infants, which have been shown to affect the relationship of mothers and their children (e.g. Muller-Nix *et al.*, 2004; Korja *et al.*, 2011; Schermann-Eizirik *et al.*, 1997). The effect of these variables will first be discussed in relation to LENA-provided information about speaker segment durations, and secondly, in relation to information gained using LENA core measures.

In the current study, LENA-detected female adult talk was found to be increased in the presence of twins, who were cared at bedside and twins who were born with normal birth weight, when compared with twins cared at a neonatal ward and twins with low birth weight. These differences were evident in the pooled data, but also in monthly comparisons. Female adults spoke more to children with normal birth weight in six of seven recordings, although the difference reached statistical significance in monthly inspections only at 10-month recordings. In addition, female adults were found to talk more to children cared at bedside in five of seven recordings, when compared with children cared at a neonatal ward. The results could be in accordance with previous research, stating that the mothers of twins would unconsciously prefer the healthier twin over the sicker twin (Mann, 1992). However, no statistically significant differences were detected between children differing for pretermity, nor between children born with and without complications.

Previously, however, the stress of having a preterm and sick baby has been shown to affect both the mother and the father (Candelori, Trumello, Babore, Keren & Romanelli, 2015). Therefore, the finding that the level of neonatal care divided groups based on the amount of male adult talk is in accordance with previous studies. Besides adult talk, it was noted that the amount of other child talk also seemed to differ between children cared at bedside and children cared at a neonatal ward. This difference was evident both in monthly data and in the pooled data. Unfortunately, the author is not aware of any studies that would have inspected the relation of neonatal health or neonatal care of younger siblings and the behavior of older siblings.

Although birth weight had importance on the amount of female adult talk in the current data, no statistically significant differences were detected in male adult and other child talk between children of low birth weight and children of normal birth weight. However, although not evident for males, other child talk in this data increased within close proximity of full-term twins, when compared with the amount of other child talk in close proximity of preterm twins.

When group comparisons were conducted with AWC and CTC, the amount of adult words and conversational turns were not found to be statistically significantly different in the pooled data between preterm and FT twins, nor between twins born with and without complications. However, in both AWC and CTC, statistically significant differences were found between the recordings of low and normal birth weight children and the recordings of children cared at bedside and children cared at a neonatal ward. In both cases, the amount was higher for children cared at bedside and children born with normal birth weight.

Although several statistically significant group differences were found in a quantified language environment, it needs to be noted that all the effect sizes remained fairly small in the pooled data comparisons. This would suggest that the effects of neonatal health-related background factors did not play a crucial role in the quantity of speech present in family homes. Additionally, it needs to be underlined again that no generalizable conclusions can be drawn from such a small sample. Instead, it could be highly likely that some of the results could turn out to be different with larger data.

When the pooled segment durations from 12-hour adjusted recordings were analyzed, mothers and key children were found to talk and vocalize slightly more on average than other children and substantially more than male adults. For male adults, the evident explanation could be that the mothers are most often the primary caretakers at family homes in the early stages of children's lives. In addition, the total duration of other child and male adult segments might also represent the range of spatial distance between family members and twin infants, as the recording devices can capture sound from an approximately two-meter radius (Oller *et al.*, 2010). However, the quantity of other child speech and/or vocal daily durations were greatly influenced by the older sibling living in a family language environment.

The effect of older siblings was evident in two ways. Firstly and very expectedly, twins with siblings heard substantially more speech from other children than first-born twins did; in the pooled data, this was supported by the substantial effect size. However, secondly and perhaps not so obviously, male adult segment durations were also found to be increased in families with older siblings, when compared with families without older siblings, although the effect size remained lower in the pooled data. These findings were also evident in all of the monthly recordings, although the differences between male adult speech quantities in families with and without siblings reached statistical significance only at 6- and 8-month recordings. This finding could, however, suggest that the presence of the older sibling might demand the father to be more active in verbal communication either with the twins or with the older sibling. The latter suggestion seems more plausible, as there was no statistically significant difference in the pooled data between CTCs from families with and without older siblings.

As for other shared demographic variables, the comparisons conducted with the fathers' educational levels did not reveal any statistically significant differences on speaker durations in any of the speaker categories. On the contrary, the most constant background variable to affect the language environment with substantial effect sizes was the mother's education. In this data, mothers with lower degrees were found to be more voluble than mothers with higher degrees, which differs from the population-level knowledge of connections between socioeconomic status, maternal education, and the quantity of spoken input to a child (e.g. Dollaghan *et al.*, 1999; Hart & Risley, 1995; Rowe, 2008). However, the children and the fathers from families with lower-educated mothers were found to be more voluble than the children and fathers from families with the mother having a higher education. In addition, CTCs were higher in the pooled data in families, where the mother had a lower education.

The abovementioned results of the relations of volubility and education are interesting, but may simply be explained by the small sample size, possibly accompanied with the fact that the division between the two groups presents only a minute difference. In Finland, both Bachelor's and Master's degrees are obtained from a higher educational system, and thus, nine out of ten participating mothers were educated for at least 15 years, and all participating mothers were educated at least 12 years. This is well above the national figure of mean years of schooling, which, in the year 2012, was 10.3 years in adults aged over 25 years (UNDP, 2013).

From the comparisons conducted with the two individual demographic variables of birth order and gender, no statistically significant group differences were found between twin As and Bs, suggesting that birth order did not influence the amount of speech and conversational turns spoken in close proximity to the twins. Gender of the twins did, however, show statistically significant differences in the amount of speech. In the pooled data, mothers addressed more speech to boy twins, while twin girls heard more other child speech or other child vocals than twin boys. There was no statistically significant group difference in the amount of speech from male adults to boy and girl twins. These findings differ from the results of VanDam (2012; referred in VanDam & Tully, 2016), who from the US-English data found that fathers talked more to sons than daughters, and mothers talked a similar amount to sons and daughters.

However, despite the results from VanDam's (2012) study, VanDam and Tully (2016) reported from a large-scale LENA study that mothers and sons were found to engage in more conversations than mothers and daughters, but the amount of CTCs did not differ between fathers' talk to girls and boys. Although the present study did not inspect the participants joined in conversations, it was found that when measured with LENA CTC, the boys were engaged in more conversations than girls,. This phenomenon was evident in both the pooled data and in six out of seven monthly recordings, although the difference reached statistical significance only in three recordings.

### 6.1.3 Twins' language development

The current study took advantage of measures used widely in clinical practices by Finnish psychologists and speech and language therapists in the hope of building up a preliminary reference base for clinicians meeting toddler-aged twins. The comparisons of twins' scores with normative information of the measures showed that twins started reduplicative babbling on average a few weeks later than children in the normative sample, but the onset of variegated babbling was substantially delayed when compared with normative information (Lyytinen *et al.*, 2000) and with the results reported by Lyytinen and colleagues (1996). The comparison of the emergence of twin's protowords was unfortunately difficult due imprecise information presented in the manual of Checklist for Vocal Development (CDEV), as the author failed to find any additional reports to accompany the information available in the CDEV manual. However, when the emergence of twins' protowords was compared with the visual presentation available in the CDEV manual, the median onset of protowords did not seem to differ substantially from the emergence of protowords in the normative data (see Graph 1 in Lyytinen *et al.*, 2000).

Thus, the possible delay in the onset of variegated babbling does not seem to affect the emergence of protowords. This can be explained by the fact that many of the first words of Finnish children have a simple geminate structure, which only requires one transition from vowel to consonant and back to vowel (e.g. "ättä" for äiti (mother) (see further Savinainen-Makkonen, 2013; Saaristo-Helin, Kunnari & Savinainen-Makkonen, 2011). However, the finding of the delay in variegated babbling is intriguing, as phonological problems have been identified as a special feature of disturbed language in twins (e.g. Bishop & Bishop, 1998; Hua & Dodd, 2000). It has been suggested that the phonological processes evident in toddlerhood could be traced back to phonetic tendencies present in the prelinguistic period (Vihman, Ferguson & Elbert, 1986).

In this study, the majority of twins seemed to have had a vocabulary spurt (see e.g. Barrett, 1995: 363, review) between the ages of 1.6 and 2.0, as the raw count of expressed words accumulated greatly during that time. However, the mean and median vocabulary and language scores remained slightly lower than the mean scores presented in the Finnish Mac-Arthur-Bates Communicative Development Inventories (MB-CDI) manual (Lyytinen, 1999). Lower scores on twin children's vocabulary are in accordance with the results from Rutter el al. (2003) study, where twins were discovered to be 1.8 months behind in development at the age of 20 months. However, the difference in this study was diminished to some extent, as the mean and median scores of the MB-CDI vocabulary and language measures approached the mean and median scores were reported in the norms at 24 months. This differs from the study of Rutter and colleagues (2003), who reported a trend of a growing difference between twins' and singletons' MB-CDI scores at 20 and 30 months.

On average, twins showed skills for the first signs of understanding and non-verbal communicative abilities at 12 months, similarly to children in the normative data presented by Lyytinen (1999). However, the mean and median scores of referential language (use of words), morphological skills, and syntactic complexity (MSL) remained lower for twins at 18 months and also slightly lower at the age of 24 months of age, when compared with normative data and data from Stolt and colleagues (2009b). In previous literature, perinatal hardships have been suggested to account for the language delay in twins (Stormswold, 2006, review), and in a Finnish study comparing full-term and preterm very low birth weight children, a difference in MSL was discovered in favor of full-term children (Stolt *et al.*, 2009b). In this study, the level of neonatal care was chosen to present the early health of twins, as the use of Apgar-scores was not feasible due to missing values. This study did not find any statistically significant group differences in any of the CDEV and MB-CDI measures between children cared at bedside and children cared at a neonatal ward. There was a small but significant difference in the mean key child segment durations in the pooled LENA data, which suggested that children cared at bedside were not as voluble as children cared at a neonatal ward. However, the difference did not reach significance in any of the monthly points of measure, and the effect size for the pooled difference also remained small. In addition, there were no statistically significant differences in monthly, nor the pooled data in the number of children's speech-like vocalizations.

The evident explanation for the above presented results may lie in the fact that all the twins participating in this study were born reasonably healthy. The children in this study did not have any severe health-related problems during their perinatal stages, nor did they suffer from any severe complications during birth and could therefore be argued to be good candidates in studying the effect of twinship in particular. However, parents did report minor complications, which had occurred during birth, and therefore, group comparisons were conducted with children, who were born with and without complications. No differences were detected in LENA key child segment durations, LENA child vocalization counts, and the vocal milestones between children born with and without complications.

Previously, Stormswold (2006, review) has concluded that non-shared perinatal complications would almost certainly account for at least some of the linguistic variation present in monozygotic twins and, in fact, statistically significant differences were detected between some of the MB-CDI scores of children born with and without birth complications. In this study, children born without complications were found to be more advanced in their early understanding and to have larger receptive vocabularies at 12 months. In addition, children born without complications were found to be more advanced in their referential use of language at 24 months, when compared with children born with complications. However, no other statistically significant group differences were detected between children born with and without complications.

There are not many studies that have reported on the emergence of variegated babbling in preterm and/or children of low birth weight, although some literature is available about the onset of reduplicated babbling (see Törölä *et al.*, 2012, review). In a Finnish study, no difference was found between the onset of canonical and variegated babbling in preterm extremely low birth weight (ELWB) children and full-term children (Törölä *et al.*, 2012). The current study also failed to find statistically significant differences between the onset of reduplicative babbling of preterm and full-term (FT) children, nor low birth weight (LBW) and normal birth weight (NBW) children. Unfortunately, it was not feasible to report group comparisons of the onset of variegated babbling between groups differing in pretermity and birth weight, as the groups remained uneven due to multiple missing values. However, interested readers may find it informative to inspect Appendix 16 for the information on group comparisons in the restricted sample of groups of twins differing in pretermity and twins differing in birth weight.

The differences in the median onsets of variegated babbling indicate that in this study (and with the available data), preterm and LBW children started their variegated babbling considerably later than FT and NBW twins. Interestingly, however, preterm infants were found to be more voluble and to produce more speech-like vocalizations than FT children. This difference was evident from nine months onwards, although the difference reached statistical significance only at 10- and 11-month measures of segment durations and at 11 months of CVC.

As for the emergence of protowords, a statistically significant result was found in groups differing in birth weight. LBW children were reported to produce protowords before NBW children, although the median difference was only a few weeks between the two groups. This finding is similar to the findings of Oller and colleagues (1998), who also reported preterm children's first words to appear earlier than FT children's words. However, the results differ from a study conducted with Finnish children; Törölä and colleagues (2012) found Finnish ELBW preterms to produce their first words later than the FT controls.

In addition to the emergence of first protowords, the group comparisons between LBW and NBW children's vocabulary development showed a statistically significant group difference in the amount of expressive vocabularies between LBW and NBW children at the age of 12 months, but failed to find statistically significant results between the size of receptive vocabulary at 12 months and the size of expressive vocabularies at 18 and 24 months of age. In this study, LBW twins had more words in their expressive vocabularies at 12 months than NBW children. In addition, preterm children had larger expressive vocabularies at 18 and 24 months, when compared with FT children, although the difference reached significance only at 18 months of age. These results differ from previous studies conducted with Finnish children (Jansson-Verkasalo, Valkama, Vainionpää, Pääkkö, Ilkko & Lehtihalmes, 2004; Stolt *et al.*, 2009b). The study of Stolt and her colleagues (2009), who reported VLBW Finnish children to acquire a receptive lexicon at a slower rate, when compared with FT children, found no statistically significant group differences between neurologically healthy VLBW children's and FT children's expressive lexicon size. It is to be noted that Stolt and colleagues (2007) did find differences in the composition of lexicon between VLBW and FT children, but this was unfortunately not examined in the current study. However, the composition of children's lexicons could also be of interest for twins in future studies.

Besides the differences in vocabularies, preterm twins – but not low birth weight twins – also had more advanced morphological skills at 18 and 24 months, although no statistically significant group differences were found between preterm LBW and full-term NBW twins in non-verbal communication at 12 months, nor in referential use of language and MSL at 18 and 24 months. However, these results do not support the suggestion that late preterms should also be considered a risk group (McGowan *et al.*, 2011).

Interestingly, the abovementioned results show that the comparisons between groups differing in birth weight and pretermity did not always produce similar results. This might underline the fact that not all LBW children are automatically preterm, nor are all preterm children automatically low in birth weight. Therefore, they should perhaps not be treated as one group, although for research purposes, it might seem plausible to treat them as one group in order to enhance sample sizes for comparative studies, especially in minor language areas.

In previous research, the role of pretermity and birth weight has been suggested to mediate with medical complications at birth (Miceli, Goeke-Morey, Whitman, Sipes Kolberg, Miller-Loncar & White, 2000). Although there were evident group differences in expressive vocabularies of children differing in birth weight and pretermity, only the children, who were born through complicated births, possessed a statistically significantly smaller number of words at 12 months. In addition, in this data, the children born with birth complications showed a tendency for smaller expressive vocabularies, although the difference did not reach statistical significance. Similarly, but not statistically significantly, children cared at a neonatal ward had smaller expressive vocabularies at the ages of 18 and 24 months, when compared with children cared at bedside.

As the early receptive vocabulary development was lower in children born through complicated births, so were the scores from MB-CDI first signs of understanding at 12 months, when compared with children born without birth complications. There were no statistically significant group differences between children born with and without birth complications in the sum variable of actions and gestures at 12 months, nor in any of the language variables (use of words, sum of morphology, or MSL) at 18 months of age. However, there was a statistically significant median difference evident at 24 months of age in the scores of word use in favor of the children born without birth complications. Besides complications at birth, no other neonatal health-related background variable was associated with group differences in the first signs of understanding or non-verbal communicative gestures at 12 months of age. Unfortunately, to the author's best knowledge, it seems that there are no published studies specifically about the relationship between language development and prolonged births or umbilical cord and breech deliveries.

In addition to the four neonatal health-related background variables, this study included five demographic background variables previously reported to be of importance in developmental studies. From these, however, the birth order of twins could be discussed both from the neonatal health and from the demographic point of view. On one hand, the birth order defines the child's role with his or her siblings, but on the other, the second born twin B has been reported to be at an increased risk of health problems and infant mortality, when compared with first-born A-twins (e.g. Purho *et al.*, 2008; Smith *et al.*, 2007; Smith, Shah, White, Pell & Dobbie, 2005).

In this study, birth order showed statistically significant group difference only in the median onset of reduplicated babbling, which for A-twins emerged earlier. No other statistically significant group differences were detected in any of the automated measures or parental report measures between first born A-twins and second born B-twins. The earlier onset of reduplicated babbling was present in eight out of ten twin pairs. In addition, in all but one pair out of eight, the onset of variegated babbling was found earlier for A-twins, while the onset of variegated babbling emerged later or was not detected during the follow-up up to the corrected age of 12 months. Although the sample was small, the difference in the onset of vocal milestones is an intriguing finding, which should be studied with larger samples. With this data, we can only hypothesize about the explanations, including, for example, the possible neonatal health-related differences in birth order not evident through the variable of birth complications or the level of neonatal care (see e.g. Purho *et al.*, 2008 and Smith *et al.*, 2007, for neonatal health outcomes in A- and B-twins) and possible differential parental treatment and attention related to first-born A- and second-born B-twins (see e.g. Lytton & Gallagher for differential treatment and Minde *et al.*, 1990, for maternal preference). However, no statistically significant group differences were detected in protowords or any of the MB-CDI vocabulary and language variables between A- and B-twins.

In previous research, the female advantage in language development has also been reported for twins (e.g. Garitte, Almodovar, Benjamin & Canhao, 2002). In this study, gender produced small, but statistically significant group differences for key child volubility and count of speech-like vocalizations in the pooled data. The pooled difference suggests that the small difference between boys and girls in this data accumulated during the follow-ups in favor of twin girls. However, no statistically significant differences were found between girls and boys in the onset of vocal milestones, the amount of vocabulary, or early language skills at any age.

Previously, maternal education has been shown to be associated with children's language development (McGillion *et al.*, 2016). Greenwood and colleagues (2011) found a tendency, but not statistically significant, for children of higher-educated mothers to be more voluble than children from lower-educated mothers, when measured with LENA child vocalization count (CVC). In this study, children of lower-educated mothers were more voluble, but the amount of child vocalizations did not show any group differences between twins in monthly inspections, nor in the pooled data. In addition, the mother's educational attainment showed statistically significant group differences in the emergence of protowords and in children's abilities to use non-verbal communicative gestures at 12 months. Children of higher-educated mothers were detected to use protowords earlier, but to have less skills in non-verbal communication. However, no other differences were detected for other MB-CDI measures.

The father's educational level did not account for group differences in twin children's volubility in monthly inspections, nor in comparisons in the pooled data or in the majority of CVC counts in monthly recordings. However, group comparisons conducted of the father's level of educational attainment did produce statistically significant group differences of twin children's expressive and receptive vocabularies at 12 months and of twins' referential use of language at 24 months. In this data, twins from higher-educated fathers had more words in their productive and receptive vocabularies, and they were more skilled in referential use of language at the age of two years, when compared with children from lower-educated fathers. In addition, the tendency was present in 24-month vocabularies and the use of referential language at 18 months, but the difference did not reach statistical significance.

In addition to the results above, children from higher-educated mothers were found to produce their first protowords earlier than children from lower-educated mothers. This finding is very much in accordance with previous studies, suggesting that SES would be correlated with twins' language skills (Thorpe, Rutter & Greenwood, 2003). However, the findings of receptive and expressive vocabularies during the second year of children's lives did not show statistically significant group differences in favor of children of higher-educated mothers. Instead, there was a tendency of children from lower-educated mothers to have larger vocabularies. This difference, although not statistically significant, was present at 12-month receptive and 18 and 24-month expressive vocabularies.

Although the differences in the amount of vocabularies are true in this data and using these methods, they could also be explained by the fact that all parents in this study were actually well-educated, and thus, the division of lower and higher education may not be appropriate. On the other hand, the finding could be in accordance with the findings of Feldman and colleagues (2000), who reported lower SES children to have larger vocabularies in parent-filled MB-CDI forms. Contrary to the mother's education, the level of the father's education showed to bear statistical significance in children's 12-month receptive and expressive vocabularies, in favor of children from higher-educated fathers. There was also a tendency, but not statistically significant, for a similar advantage of later 18- and 24-month vocabularies of children in this data.

Although siblings are often thought to negatively influence a child's language acquisition by reducing the amount of child-directed speech (CDS) from mother to child (Oshima-Takane et. al., 1996; Tomasello & Mannle, 1985; Woollet, 1986), there are also suggestions that having an older sibling could be beneficial to learning (Brody, 2004; Barr & Hayne, 2010; Hoff, 2006). In addition, Shneidman Arroyo, Levine and Goldin-Meadow (2013) have previously stated that children with siblings actually hear a similar amount of CDS to first-born children, but when overheard speech is accounted for, children with siblings hear more word types and tokens than first-borns. In a study by Haapsamo and colleagues (2013), siblings were found to enhance family relationships, but not to have an effect on a younger sibling's vocabulary development. In this study, several differences were detected between twins with and without siblings: 1) twins with siblings were less voluble and vocalized less during the recordings, but 2) they were more advanced in non-verbal communicative gestures at 12 months, 3) had larger vocabularies at 24 months, and 4) had more advanced morphosyntactic skills at the age of 24 months, when compared with first-born twins. Previously, it has been shown that early symbolic gestures are beneficial to children's verbal development (Goodwyn, Acredolo & Brown, 2000). These findings are intriguing, as they might imply that older siblings could in fact enhance a younger sibling's development – at least in families, where parents face the demand of sharing their attention between the two twin children (for a positive effect of siblings on children with Autism, see Ben-Itzchak, Zukerman & Zachor, 2016). Thus, the results from this study are in accordance with Barr & Hayne (2010) and Brody (2004), who stated that younger siblings can learn from older siblings.

### 6.1.4 The associations between language environment variables and child language development variables

As the current study employed a two-fold perspective to investigate questions related to the methods and questions related to information gathered using the selected methods, both aspects are inevitably present when discussing the relations of information gathered and analyzed in this study. Therefore, this chapter first discusses the associations of information derived using LENA and parental methods. The relation of quantity in twin children's language environment and the language outcome. Secondly, the current chapter discusses the concurrent validity of the LENA System in comparison to more traditional measures. Additionally, the second part inspects the relations within these instruments (intra-correlations) and compares the internal correlations to information available about the intra-correlations of CDEV and MB-CDI measures (see also 2.3.2.1 and 2.3.2.2)

This study found moderate positive associations between female and male adult segment durations, but not between the durations of adult speakers and child volubilities. If we assume LENA to work properly, these results suggest that the amount of adult talk did not increase or decrease the vocalization activity of the twins in this data. Similar interpretation can be presented from correlations within LENA core measures. In this study, LENA AWC and CTC were found to correlate, as did CVC and CTC, but no statistically significant association was found between AWC and CVC. Previously, similar results have also been presented with American-English data (Greenwood *et al.*, 2011).

In the past research, the amount of parent speech at 16 months has been suggested to relate with vocabulary growth (Huttenlocher *et al.*, 1999) and maternal speech to 18-month-olds to correlate with language proficiency at 24 months (Hurtado *et al.*, 2008). The associations between information gathered with automated methods and parental questionnaires in this study were inspected only from the pooled data, instead of monthly points of measure. None of the speakers and none of the core measures showed concurrent associations within CDEV and MB-CDI. However, single statistically significant associations were detected. 1) The increased amount of male adult talk was associated with later onset (or detection) of variegated babbling and with more advanced use of non-verbal communicative actions and gestures. 2) The increased amount of female adult talk was associated with later emergence (or detection) of protowords and, as with male talk, with more advanced use of non-verbal communicative actions and gestures. In addition, 3) the increased amount of key child volubility was associated with smaller receptive vocabulary and lower scores on the first signs of understanding at 12 months.

However, the increased amount of male adult talk was found to be positively associated with the size of vocabulary at 24 months, and the amount of other child speech was found to be positively associated with syntactic and grammatical skills at the age of 24 months. These results could reinforce the statement suggesting that that all input is important, not just the talk from mothers (Schneidman *et al.*, 2013). However, in the majority of cases, no statistically significant associations were found between speaker segment durations and vocal milestones and vocabulary at 12, 18, and 24 months. This could suggest that the uniformity of the amount of LENA-detected speech and CDEV-measured milestones and MB-CDI-measured language skills might not be strong. But, as the current data was very limited, this should be studied with larger samples.

To the author's best knowledge, it seems that the LENA-provided segment durations have not been previously inspected in relation to outcome measures. However, there are several studies, which have reported the associations between LENA core measures and traditional clinical measures; the amount of CVC and CTC have been reported to correlate with the comprehension, expression, and total scores from the Preschool Language Assessment (PLS-4), but not from the Bayley Scales of Infant Development (BSID) (Greenwood *et al.*, 2011). In addition, AWC has been found to correlate moderately with PLS-4 scores ($r=.35$, $p<.05$) and CTC (but not AWC), to correlate with receptive language abilities of children, who are hard-of-hearing (VanDam, Ambrose & Moeller, 2012). It is to be noted that in a study by Warren and colleagues (2010), the MB-CDI score was found to be substantially correlated with CTC ($r=.80$, $p<.01$), but Ramirez-Esparza and colleagues (2014) did not find statistically significant associations between MB-CDI vocabulary and AWC.

The current study failed to find concurrent associations between LENA core measures and early language development measures with parental questionnaires. LENA-provided CVCs and AWCs did not statistically significantly correlate with any of the CDEV milestones, but the increased amount of CTCs was strongly associated with the later emergence (or detection) of protowords. The non-significant association between child vocalization counts and later development differs from previous studies (McCarthren *et al.*, 1999, review). As for the MB-CDI vocabularies, the only statistically significant association was found for AWC and 12-month productive vocabulary. the correlational analysis suggested that the more adult words the child heard, the fewer words the child had in their expressive vocabulary at 12 months. In addition, the only statistically significant correlation for linguistic variables of the MB-CDI and LENA core measures was found in AWC and non-verbal communicative gestures at 12 months. This association suggested that in this data, children, who heard more adult words, were more advanced in their non-verbal communication at 12 months.

The results presented above do not confirm that the LENA core measures would have been able to predict language development of twins in this data, had MB-CDI been used as an outcome measure. However, the present results do not imply that the environment surrounding the child would not be meaningful, but the current study suggests that more work is needed in order to study the predictive validity of LENA measures. In addition, as the heavy weight of the research has underlined the importance of child-directed speech (CDS) on first language acquisition (Golinkoff *et al.*, 2015; review; Soderstrom, 2007, review; Weisleder & Fernald, 2013), automated detection of the CDS from language environment could be of great value to researchers and clinicians. For this, an easy way of collecting the data with LENA DLP could be beneficial, although the LENA program can only account for the raw quantity of speech. However, to the author's best knowledge, it seems that there are currently no other well-working solutions for automated detection of CDS (see, however, Vosughi & Roy, 2012, for CDS detection in the Human Speechome project).

Canonical babbling and first words have not been found to be correlated (Oller *et al.*, 1998), but the onset in variegated babbling has been shown to be related to word production (Keren-Portnoy, Majorano & Vihman, 2009). In a recent study, babble was reported to predict 22 percent of the variance of the onset of first words (McGillion *et al.*, in press). The current study, however, failed to find statistically significant associations within parent-reported CDEV vocal milestones. The onset of reduplicated babbling, variegated babbling, and the emergence of protowords were not associated to each other in this sample of twins. In addition, no statistically significant associations were found with CDEV vocal milestones and MB-CDI vocabulary and language scores. These results suggest that the onset of reduplicated babbling, the onset of variegated babbling, and the emergence of protowords in twins in this study were not associated with 1) each other, nor 2) the vocabulary, and 3) language skills during twins' second year of life.

Previously, a study of the validity and the predictive power of the CDEV in relation to MB-CDI found that children who achieved prelinguistic milestones early were also more competent in their language proficiency at the age of 18 months (Lyytinen *et al.*, 1996). However, although this study did not find similar connections as Lyytinen and colleagues (1996), the results from the present study are in accordance with the information presented in the CDEV manual (Lyytinen *et al.*, 2000). The CDEV scores were not found to be associated with MB-CDI vocabulary and language measures at 12, 18, and 24 months, although statistically significant but moderate correlations were reported between CDEV and MB-CDI at 14 and 30 months.

From the preliminary inspections of the predictive validity of the Finnish MB-CDI, Lyytinen and colleagues (1996) reported that MB-CDI and RDLS were correlated, and that higher uniformity was found for MB-CDI productive measures. In the final product of the normative study – the method manual – the amount of receptive and productive vocabularies was, however, reported to have only a weak to non-existent associations with Bayley MDI and with RDLS (Lyytinen, 1999). Additionally, no associations were reported between MB-CDI 12-month nonverbal communicative gestures and Bayley MDI at 24 and RDLS understanding or production at 30 months (Lyytinen, 1999). This finding was replicated with full-term children in a study of Stolt and her colleagues (2016). However, although there were no associations with full-term children, Stolt and colleagues found the MB-CDI gestures to correlate with child language skills at the age of five years in VLBW Finnish children, with and without known neurological impairments (Stolt *et al.*, 2016).

The low predictive power of MB-CDI 12-month evaluation has been raised for discussion by Feldman *et al.* (2000) and acknowledged by Fenson, Bates, Dale, Goodman, Reznick, and Thal (2000). However, in this study, the 12-month receptive (but not productive) vocabulary showed to be positively correlated with the referential use of language at 18 and 24 months and with MSL at 24 months. In addition, in this study, vocabulary and referential use of words at 18 months were associated with MSL at 24 months and morphological scores at 18 months were associated with morphological scores and MSL at 24 months. However, as the sample in this study was very limited, it needs to be remembered that the predictive power of MB-CDI at 18 months has been previously questioned (Duff, Nation, Plunkett & Bishop, 2015).

Although the predictive validity of applying MB-CDI to the earlier stages has shown to be questionable, the use of referential language, morphological skills, MSL, and the amount of productive vocabularies at 18 and 24 months were found to be moderately correlated with RDLS understanding and production at 30 months in the normative Finnish study (Lyytinen, 1999). In addition, there are studies reporting the predictive value of vocabulary from two years even up to 13 years of age (Rescorla, 2005). In one Finnish study, the amount of productive vocabulary at the age of two years has been shown to predict language comprehension at three years (Korpilahti, Kaljonen & Jansson-Verkasalo, 2016) and to associate with picture-naming abilities at the age of five years (Vainio, Haataja, Lapinleimu, Lehtonen, Stolt & PIPARI-study group, 2011). Additionally, the lower scores in MSL at the age of two years have been reported to be different between children at familial risk of dyslexia and without the familial risk (Lyytinen, Poikkeus, Laakso, Eklund & Lyytinen, 2001).

Besides the evaluation of concurrent validity and the predictive evaluation of the MB-CDI, the internal correlations are also important, as they provide information about the consistency of the instrument. Previously, internal validity of MB-CDI was reported to be satisfactory, and earlier measures have been presented to predict later measures, accounting for 16.6 to 31.1 percent of variance (Bates, Dale & Thal, 1995: 101). Additionally, the MB-CDI within measure correlations have yielded substantial associations for vocabulary, MSL, and morphological complexity in the Finnish MB-CDI method manual (Lyytinen, 1999). In the current study, many of the MB-CDI measures were found to be substantially correlated with each other, with the highest amount of associations present in earlier measures and MSL and vocabulary at 24 months. However, multiple significant connections were also found with the 18-month receptive and productive vocabulary, as well as with the 12-month receptive vocabulary. On the contrary, in this study, the 12-month productive vocabulary did not correlate with any of the other measures in any points of measure and the first signs of understanding were correlated only with the 12-month receptive vocabulary.

## 6.2 Methodological considerations

In the age of digitalization, the distinction to identify the relationship of science and technology and to ponder on the material realization and theoretical interpretations have been emphasized as important subjects for the study of scientific experimentations (Radder, 2003: 7–8; Radder, 2009). This has been acknowledged in the field of child language, for example, by Dan Slobin, who has stated that “the history of science is as much a history of discoveries, as it is a history of the tools that make discoveries possible” (Slobin, 2014: 1). The current study gained information through traditional parental questionnaires, large-scale recordings, and automated analyses made possible by a novel LENA method. Besides analyzing the information about child language and twins’ environment, the novel method was studied first, applied second, and thirdly, the information gained from the novel method was compared to information from more traditional methods. Although the current work could not validate the LENA method, the selection of the two-way approach was an attempt to take into account the fact that all empiricist information is always dependent of the ways the data is collected. With the rise of automated techniques, a critical evaluation of our expectations for technology is needed, as well as discussions related to the questions of the justifiability of the results of the new approaches (Radder, 2009).

### 6.2.1 Study design, data collection, and sample characteristics

This study was designed to gather information about twin children's early vocal and later language development and to acquire information about the natural language environment in twin families. As this study was conducted from the logopedic viewpoint, the theoretical background and research design was built acknowledging the diversity and complexity of influencing factors relevant to language acquisition. In this study, the complexity of the multiple possible factors behind the studied behavior was emphasized, although the studied group remained small and hence the results can only be considered to describe the studied group of twins, instead of representing all twins. In addition, this decision lead to large number of studied variables, which inevitably influenced the clarity and depth of the present work.

In data collection, the feasibility of data collection for participating families was a key issue, as families were already facing the demanding task of taking care of two preverbal babies. This meant that all data needed to be acquired without difficult devices, participant travelling, or other significant efforts from the families. From this viewpoint, the design was successful, as several parents reported that the participation in the study felt effortless and home-delivered recording devices were a pleasant way to participate in research. In addition, all families were committed to the study for the whole time, although the data collection lasted for a total of 1.5 years per family.

As the study was originally designed to be completed within the four-year period, no strict sample size could be determined. Thus, as the interest was, in this case, in acquiring information about the developmental process, the sample size in the present study remained small. The number of participants could have increased, if recruitments would have been conducted systematically using register information to recruit participants or the design would have been cross-sectional. However, all the recruitments were based on volunteer family activity to make contact with the researcher. Besides sample size, the selected procedure in recruitment also possibly influenced the composition of the sample, as all the parents in participating families were well-educated and seemed to have interest in their children's development.

In this study, audio data was recorded during the two first years of the four-year study period and last follow-ups were conducted three years after the onset of data collection. Although the sample remained small and not representative in regards to the entire Finnish twin population, the audio samples can be argued to be more representative due to longer recordings, when compared to samples in previous research (Ambridge & Rowland, 2013). Yet the recordings were conducted only monthly, which may not be a dense enough sampling in the era of rapid developmental changes (e.g. Lieven & Behrens, 2012). For example, Adolph and colleagues (2008) found that the follow-up of developmental trajectories of motor development lost sensitivity for variation, if the intervals between observations were longer than seven days. Thus, this study continued the tradition of "painting snapshot portraits", as critiqued by Adolph and colleagues (2008). Even if this is the case, this study did produce a small but novel amount of information about the prelexical development of twins with the methods developed for clinicians.

As the recordings were conducted monthly, the researcher had the opportunity to control and remind families about the importance of filling out the Checklist for Vocal development (CDEV) regularly to ensure delivery of reliable information of the times when children acquired new skills. After the recordings, the MacArthur-Bates Communicative Inventories (MB-CDI) forms and return envelopes were posted to families at selected measuring points with the request to fill out the forms as rapidly as possible. Because there were only ten families, controlling the use of the MB-CDI forms was an effortless task; the researcher called parents, if the MB-CDI forms were not returned within a two weeks' time after posting. This procedure confirmed that parents did the evaluations of their children's lexical and grammatical skills.

Besides the easiness of data collection for the families, the selection of measurement tools was based on clinical applicability, reliability of measures in previous research, and the interest in novel automated approaches. The decision to apply LENA analyses was not, however, an easy one as the validation of the system was beyond the scope of this study, and twins can be viewed as special population. If the LENA System was to be validated for Finnish, the sample should also contain singletons and singletons with siblings. Therefore, all LENA analyses needed to be interpreted and generalized with great caution. A few words should also be said about the sample and the way LENA reliability was studied, as the decisions in some parts differed from previous research.

Firstly, in this study, LENA reliability was inspected with two full-day recordings, instead of randomly selecting an arbitrary size sample from the 1500-hour data collected from twins and their immediate environment. A disadvantage of this decision lies in the fact that it is possible that the two selected recordings were not from a typical day from a typical family. Instead, it is possible that the recordings were from families, whose LENA analyses would be unreliable due to reasons related to family interaction, the acoustic properties of the home environment, or family members' speech. All these types of sources of error would be diminished in more representative data. On the other hand, the advantage of the selection was that by analyzing whole-day recordings, it was possible to get to know the data and to gain experience from possible error sources in LENA transcriptions. In addition, the two recording samples were a total of 22 hours, which exceeds the samples analysed in the studies evaluating LENA performance for Chinese and Spanish, but not for French (Canault *et al.*, 2015; Gilkerson *et al.*, 2014; Weisleder & Fernald, 2013).

Secondly, this study used pre-segmentations from LENA, and thus, agreement between humans and LENA were higher than in previous reports. However, LENA reliability has so far been studied mostly with agreement rates for segmentation and inter-rater correlations for LENA CVC and AWC core measures, but this study tried to apply a more versatile way of exploring reliability. That said, it needs to be acknowledged that some studies have reported Cohen's kappa (e.g. VanDam & Silbert, 2013a), reports on sensitivity and precision (Gilkerson *et al.*, 2014), or the number of false positives and negatives (Oller *et al.*, 2010) in addition to the percent agreement. Furthermore, the absolute errors are presented in addition to correlation coefficient (e.g. Soderstrom & Wittebolle, 2013; Canault *et al.*, 2015). The use of the absolute error (or relative error and percentage error) informs the reader more accurately about the possible significance of the error or, if presented from multiple samples, informs the reader about the nature and magnitude of error and error variance. This is a benefit when compared to correlational analyses, which only present the strength of the association between the observed scores, thus, leading to multiple potential mistakes in interpretation (Bland & Altman, 1986; 1999). As the evidence-base of LENA reliability and validity is only building up, the need for precise information about its performance should be emphasized.

Additionally, the derivation of samples and the determination of sample sizes and lengths should be considered and discussed in relation to the information presented in LENA technical reports. In this study, LENA and human counts were compared hourly. This selection was based on information from LENA technical reports. Previously published studies have used a variety of ways of obtaining samples, although in LENA, the appropriate length of the samples makes sure that the statistical models work properly (see Xu *et al.*, 2008; LTR-05-02, reliability over time). In one of the LENA technical reports (LTR-05-02; Xu *et al.*, 2009), the error of LENA AWC estimates is reported to decrease to acceptable rates after approximately one hour, and a similar result also was found by Canault and colleagues (2015).

### 6.2.2 Strengths and weaknesses of data analysis

In the current study, the <u>reliability of the LENA System</u> was analyzed using human the gold standard, as has been common practice in previous LENA reliability studies (e.g. Xu, Yapanel & Gray, 2008; Oller *et al.*, 2011; VanDam & Silbert, 2013a). In addition, the study included an evaluation of the reliability of the human coder with a human inter-rater procedure. This procedure yielded good agreement between primary and secondary raters. However, the agreement remained lower between human inter-raters when compared with human-LENA agreement. This finding could be interpreted in different ways. It is possible that the primary coder was in high agreement with LENA because the human listened to the recordings thoroughly, and thus, was more acquainted with the voices of the speakers and was able to use contextual information to interpret sounds, therefore identifying speakers more accurately than her inter-raters. However, it is also possible that the primary coder was somewhat more influenced by the pre-segmentation than the human inter-raters. In addition, the researcher acknowledges that the pre-segmentation could above all affect the identification of twin siblings from each other. In future studies, the problem of misclassifying children on the same developmental level could be avoided, for example, by studying the agreement of multiple raters and without the visible pre-labelling of the LENA System. However, the author still suggests that in such studies, the human raters should have the benefit of listening to the recording enough times to get acquainted with the children's voices and the context of the interactional situation to help separate the children from each other.

In the current study, the LENA identification of speakers was designed to include an evaluation of LENA performance in both near/clear and far/faint segments. This selection, however, did not prove to be an optimal one for the current study, as LENA segment durations were calculated only from the near/clear segments. However, in spite of this selection, most of the LENA speaker identifications turned out to be fairly reliable. In addition, the current results suggest that the LENA pre-segmentation and calculations could be used, for example, as the basis of systematic language sampling (see e.g. Petäjistö, 2015).

In addition to LENA reliability, the discussions of <u>LENA internal consistency and its convergent validity</u> in comparison with traditional measures should be of great interest to anyone interested in piloting, applying, or validating the LENA System for different linguistic and cultural areas. In this study, validity was inspected using parent reports, as they were the only standardized measures available for Finnish children under the age of one and two years. In addition, the author also recognizes that if LENA was to be validated for Finnish, more versatile assessments would be needed and with a different type of sample from the population of Finnish children.

The <u>measuring of spoken interaction in family environments</u> in this study was conducted with the LENA System. Through the LENA System, information about the volubility of different speakers was gained, but as the current study did not have comparative groups of singletons, the analysis was limited in that sense. The author acknowledges that the current work would have benefitted greatly from singleton comparisons with the LENA-measured information, but also argues that the measured quantified speech is a also valid variable on its own and in regards to correlative analyses on the relations of received input and development of measured language skills. The LENA measures, for example, showed that the vocal time of key children did not accumulate during the studied period from 6 to 12 months. This is a small but a novel piece of information, even if the comparison to singletons is lacking.

The statistical methods selected for the analyses of the meaning of background variables and for the associations between measured variables were all non-parametric in nature. Non-parametric Mann-Whitney U-test (later M-U) was selected to study the magnitude of difference between sub-groups of twins both in their language development and in the quantified speech in families. The selection of group comparisons with M-U, although technically valid, was problematic in the sense of producing a large amount of comparisons with the difficulty of interpreting the actual information value, especially for the comparisons between the amounts of spoken interaction in families.

For the group comparisons of the effect of background variables on quantified speech in family homes, the effect sizes were calculated only from the accumulated pooled data. This selection aimed to diminish the proportion of error in measurement. However, the selection also produced a new problem with interpretation. If there were no consistency in monthly group differences of selected variables, how could the small accumulated statistically significant differences be interpreted? Previously, Hart and Risley (1995) reported that the everyday differences in family interaction accumulate in time and eventually lead to large differences in the mastery of vocabulary. On the other hand, the studied group in this study did stay within normal variation in their language development, thus raising the question of whether the found group differences actually made any difference in children's language abilities.

The question of whether or not LENA-quantified spoken interaction showed to be meaningful in regards to children's language development gained an additional angle from the correlational analyses between the LENA variables and variables from parental questionnaires. In these analyses, the non-parametric Spearman rank correlation coefficient $(r_s)$ was applied. However, the correlational analyses were sometimes hard to interpret, as the correlations have dual meanings. On one hand, they can be seen to present the actual association of the intendent measured phenomena (e.g. the relations of a child's vocal activity to later expressive vocabulary), but on the other hand, correlations can be interpreted to give information about the actual usability and validity of the measurement itself (i.e., is it feasible to quantify child vocalizations, if it does not have associations with well-established measures of development). However, this question can be addressed to both LENA and the Child Vocalization Checklist (CDEV) measures, as CDEV milestones also did not show to correlate with later language development.

### 6.2.3 General discussion and ethical considerations

The theme of pursuing accurate and reliable information about children's language abilities and developmental processes unites clinical researchers – speech and language therapists – and fellow academic researchers. As the LENA System is experienced as a feasible tool for parents and easy to use for clinicians, the implementation of its use in clinical practices would be easy, and it would serve as a non-biased tool free from variation inevitably present in the evaluations made by humans. But, as the system is based on American English, the reliability of LENA should be studied before it can be applied to other language and culture areas. In addition, as there is still no consensus about the importance of and associations between spoken input quantity and quality, studies focusing on the predictive validity of the LENA measures are needed. As stated by Heilmann, Miller, and Nockerts (2010b): "In the age of accountability and least-biased assessment, it is the responsibility of clinicians and researchers alike to critically evaluate their assessment practices".

In addition, if the LENA System was to be adapted in clinical practices in Finland, it would first require a careful consideration of how to store raw data, as all recordings in clinical practices would be handled as patient documents, instead of research data, and the storage of patient data is heavily legislated in Finland. Apart from data storage challenges in health care, there are critical voices concerned about the extensive use and safety of using the LENA System with large-scale populations like the Providence Talks project (Rosen, 2013). Although the social engineering surveillance problem articulated by Christine Rosen may appear to be far from the innocent intentions of understanding and enhancing child development it seems to present, there is still a need for a thorough ethical questioning in relation to LENA and other similar tools. What type of information can be collected and for what use? Do the benefits of learning from the data exceed the potential harms and hazards that are unlikely, but possible to occur? The latter question is especially related to the analyses of the associations of LENA-quantified variables and later child development. Future studies should a at least address the questions of 1) what (if any) highly important information we gain from LENA that we should decide to record a large amount of sensitive data from people, and 2) what are the possibilities of applying the LENA (or similar) technology, and what are the possible negative effects of using the system?

Although LENA may well turn out to be a valuable tool in addition to traditional measures used in clinical practices, it cannot deliver qualitative linguistic information, which is also undoubtedly needed in the future in order to understand the language capabilities of typically developing children, healthy adults, and also different groups of speech and language therapy clients. In addition, as the buzz of health technology and computational rehabilitation method applications is rising, we should bear in mind that no tool should define what we rehabilitate or study. Instead, it should enable us to study whatever past research has guided us to study, or what our empirical findings suggest would be worthy of study. In addition, clinical and academic researchers in behavioral sciences, and specifically in logopedics, need to discuss the scope of interest in studies related to developmental processes in typical populations and rehabilitation contexts. According to Pickstone *et al.* (2009), besides asking "what works", we should also be interested in finding out the theoretical foundations of interventions ("how do interventions work"), populations that benefit from interventions ("for whom they work"), and contextual factors in implementing interventions ("in what circumstances do they work"). None of these questions can be answered with simple models, but they require understanding of different forms of input and other factors influencing the dynamic processes of learning (Golinkoff *et al.*, 2015).

As previously reported, children with language delays and impairments have been identified to more often be from low SES families or from lower-educated parents and to more often be boys than girls (Stanton-Chapman *et al.*, 2002; Korpilahti *et al.*, 2016, Wallace *et al.*, 2015). In addition, neonatal health, specifically related to birth weight, and 5-min Apgar scores have shown to be of importance to language development (Stanton-Chapman *et al.*, 2002). Some of the background variables also showed importance in this study. However, as this work acknowledges the complexity of factors influencing development, and thus, follows the emergentist, dynamic systems and transactional views, the author does not attempt to claim to be able to leash all influencing factors present in children's lives. Instead, the author sees this study as a manifestation of some of the problems faced within behavioral sciences, and specifically, within logopedics. On one hand, the researcher needs to deal with the pressure to over-simplify complex developmental processes or the clinically gathered information to acquire etiological explanations from developmental phenomena, which are actually affected by numerous intrinsic and extrinsic variables. On the other hand, the researchers and clinicians need to manage and draw conclusions as reliable as possible from the complex caseload, consisting of a limited amount of participants.

## 6.3 Applicability of the results and recommendations for future studies

This study was conducted with a small sample of reasonably healthy set of twins, and thus, the sample is not representative. However, as we do not currently have information about the prelexical development of twins (see, however, Nan *et al.*, 2013), this study gives some preliminary information about the vocal development and early language skills of twins. The information from the current study could serve as a facilitator of future studies related to twin children's early language development and the factors affecting it. In addition, the insight gained through the current study could be used in planning well-designed future studies.

As times are changing in the current era of digitalization, more and more technical advancements are also sought after in health care. Although this study did not validate the LENA System for Finnish, the small reliability trial nevertheless raised interesting points up for discussion, which should be taken into consideration, if validation of the system was to be considered. These include both the way the reliability test was constructed and its results, but also the internal and between-tests correlational analyses to measure the LENA System's validity. Besides the easily usable core measures, this study sought the total durations of the vocal activity of the speaker identified by LENA. Both the reliability of segment identification in general and the proportion of clear "meaningful" signal are important, if the segmentations are planned to be used, for example, as a basis for systematic language sampling.

Children's speech and language therapy interventions can be divided into child-focused or environment-focused approaches (Pickstone *et al.*, 2009, review), but in general, the assessments and interventions are based on the ICF-model, which emphasizes the person's ability to participate in everyday life activities and the enabling role of the environment and ecological rehabilitation (WHO, 2001; McLeod & Bleile, 2004). The LENA-provided information about the amount of close speech, interactions, initiations, and engagement in conversation could be applied in ecological interventions, such as parent-child interaction therapy (PCIT) (Falkus *et al.*, 2015), as a way of giving feedback to the family, caretakers, and other significant people in the life of the client, but also as part of an assessment of the efficacy of ecological therapy interventions in various clinical subgroups (for a pilot test on the use of LENA in intervention, see Sacks and colleagues (2014). In fact, there are already several on-going studies applying LENA in early interventions, including the large-scale 30-million-word initiative (http://thirtymillionwords.org/; Leffel & Suskind, 2013) and pilot-studies, which are expanding the use of LENA within non-English language areas (e.g. Löfkvist, 2016).

As large databases like Homebank (VanDam *et al.*, 2016) are building up both from raw data and from LENA analyses, we need to make sure that we can trust the data and the results derived from it. Currently, the most accurate information about LENA performance is available in the reports focusing on the accuracy of speaker segment durations, although the accuracy of segment boundary placement is still missing, and therefore, we do not have the information about how much is excluded from the segment duration calculations exactly. As the majority of core measure reliability studies have used correlational analyses to find out the agreement between human and LENA counts, additional studies are needed. Besides correlational analyses, future studies should further inspect the actual and relative errors made by LENA and especially the variation and consistency of the errors. In addition, we should discuss whether LENA can be seen to represent the true world and continue to ask if the information we get is actually meaningful.

To serve these goals, LENA needs to be studied with normal populations, and large enough samples need to be collected of clinical groups with significant contrasts in TDs. In addition, we need to further study the predictive and concurrent validity of LENA both in relation to standardized language assessments, but also in relation to the quality of home interaction to replicate the findings from the associations between input quantity and quality from the Hart and Risley (1995) study. If LENA turns out to give us accurate enough information, the LENA data combined with experimental techniques might give us insights into the dynamic learning processes and the relation of the environment and fundamental abilities enabling learning (see Odean, Nazareth & Pruden, 2015).

To this date, it is well-recognized that LENA DLP serves as an effortless tool for collecting data, and this needs to be appreciated, even if we observe problems in applying the LENA algorithm with non-AE data, or if future studies show that the measurement of pure quantity does not benefit children needing interventions. The LENA-provided raw data could be used in the most imaginative ways, if researchers in logopedics, psychology, and pediatrics reached towards the expertise of computational linguists, researchers from signal-processing, and related fields. Although a lot of work is already conducted on audio and speech recognition, we are still in demand, for example, of reliable CDS detectors, word and grammatical detectors, and sound to text-converters. In addition, the possibilities of acoustic signals are not yet fully utilized, but new possibilities are emerging, for example, as spatial audio analysis techniques are advancing (e.g. Pertilä *et al.*, 2013).

Besides the current LENA applications, the system provides intriguing possibilities of applying the system in clinical populations not yet studied, including post-stroke adults and people with progressive degenerative illnesses – all populations with conditions affecting a person's activity and participation, including the ability to use speech and language (see Li *et al.*, 2014). We could, for example, study the effect of heard input on spontaneous recovery, the elicitative effect of conversational turns, and combine information about quantified interaction with people's self-assessed perceptions of their quality of life and participation in everyday life events. In addition, other technical solutions might also be worth studying, for example, using tablets in data collection (Frank, Sugarman, Horowitz, Lewis & Yurovsky, 2016) and conducting computer-assisted language sample analysis in a unified manner (see e.g. Heilmann, Miller & Nockerts, 2010a).

To set technological questions aside, the main finding of this study was the possibly enhancing role of older siblings on twin children's language development and on the father's involvement. In addition, the current study also encourages further study into the father's role in children's language development, as the amount of male adult talk was associated with a larger vocabulary at the age of two years (see also Sarkadi *et al.*, 2008). Within the possibilities offered by LENA technology, it would be, for example, possible to further analyze conversational patterns and find out who the participants interacting with the key child are, and whether this involvement shows associations with later development. In addition, with the current participants, a follow-up in the later stages of children's lives would be of great interest, as it would give information about twins' later development and the possible associations between the measured environment and vocal activity from prelexical stages to the family activity and children's language proficiency in preschool and school years. However, for more generalizable information, larger samples of singletons, twin children, and their families should be also studied.

Besides the effect of an older sibling, another intriguing finding was the delay in the onset of variegated babbling in twins. As the specific feature of twins' language development is thought to be disrupted phonology (Hua & Dodd, 2000; McMahon *et al.*, 1998), future studies involving early-age follow-ups could help in clarifying the possible causes and results of the phonological development of twins. As a mother of twins herself, the author has often pondered on, for example, the possible influence of increased distance and lessened experiences of straight one-on-one experiences by the twin children: could the shortage and/or scarcity of obtaining close visual information and the possibly more often experienced extended distance between the speaker and the language-acquiring child affect the statistical learning of phonemes and phonology? To answer the question related to the effect of distance of caretakers on the language-acquiring child, LENA data could be analyzed in novel ways by applying, for example, state-of-the-art audiospatial analyses of the current data (P. Pertilä, 2015, personal communication). The techniques also show a promise for researchers interested in the experiences of the shared and non-shared social environment. By applying these novel techniques to inter-twin studies, we might gain insight into the yet unrecognized non-shared social environmental factors, which are without a doubt influencing the development of language-acquiring children every day.

# 7 Conclusions

This study inspected the reliability of the LENA measures, the language development of twins in relation to normative information, the effect of biomedical and social environment to twins' language development, and the uniformity of information as an indicator of LENA validity. The main conclusions include statements about the methods, statements about child development, and statements about the relations of social and biomedical environment and child language development.

For the LENA reliability, the main statements are: 1) LENA key and other child identification was reliable. 2) LENA female adult identification was reliable, but LENA male adult identification was less precise. 3) LENA identification of speech-like vocalizations was reliable. 4) LENA child vocalization count (CVC) was reliable with a moderate amount of underestimation, but 5) LENA adult word count (AWC) was not reliable, and the errors in word count estimates were not consistent.

For the child development, the main statements are: 1) Twins vocalized slightly more than children in LENA AE-based normative data. 2) A-twins started reduplicated babbling earlier than B-twins, but the onset of reduplicated babbling was not delayed. 3) The onset of variegated babbling was delayed in twins, when compared with normative information. 4) On average, twins scored lower in vocabulary and language than children in the normative data, but 5) they remained within normal variation.

For the social and biomedical environmental influences, the main statements are: 1) Older siblings influence family interaction by activating fathers and by reducing the time their infant siblings vocalize. 2) Twins cared at bedside heard more speech from family members than twins cared at a ward, and they were found to be less voluble. 3) Twins with older siblings had better language outcomes at the age of two years than first-born twins. 4) Mothers', but not fathers', education was associated with multiple group differences within the quantified information from family interaction.

For the uniformity of the methods, the main statements are: 1) LENA measures were not systematically and strongly associated with language outcome from parent reports. 2) LENA adult and child segment durations were not associated with each other. 3) LENA CVC and AWC were associated with CTC, but not with each other. 4) CDEV milestones were not associated with each other, nor with MB-CDI vocabulary and language. 5) MB-CDI showed internal associations between the majority of the measures, except for 12-month expressive vocabulary and the first signs of understanding.

# REFERENCES

Acarlar, F. & Johnston, J. (2009). Computer-based analysis of Turkish child language: Clinical and research applications. *Journal of Multilingual Communication Disorderds, 4*(2), 78–94.
Adolph, K.E, Robinson, S.R., Young, J.W. & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review, 115*(3), 527–543.
Adolph, K.E. & Robinson, S.R. (2011). Sampling development. *Journal of Cognition and Development, 12*(4), 411–423.
Aldosari, M., Almuslamani, A., Wilson, F. & Gilkerson, J. (2012). *Validation of language environment analysis in Arabic-speaking individuals.* Poster presented in International Meeting for Autism Research, Toronto, Canada.
Ambridge, B. & Lieven, E.V.M. (2011). *Child language acquisition: contrasting theoretical approaches.* Cambridge: University Press.
Ambridge, B. & Rowland, C. (2013) Experimental methods in studying child language acquisition. *Wiley Interdisciplinary Reviews. Cognitive Science, 2*, 1–20.
Ambrose, S., VanDam, M. & Moeller, M.P. (2014). Linguistic input, electronic media, and communication outcomes of toddlers with hearing loss. *Ear and Hearing, 35*(2), 139–147.
Anand, D., Platt, M.J., Pharoah, P.(2007) Comparative development of surviving co- twins of vanishing twin conceptions, twins and singletons. *Twin Research and Human Genetics 10, 1*, 210–215.
Aragon, M. & Yoshinaga-Itano, C. (2012). Using language environment analysis to improve outcomes for children who are deaf or hard of hearing. *Seminars in Speech & Language, 33 (4), 340–353.*
Bandura, A. (1971). *Social learning theory.* New York: General Learning Press.
Barr, R. & Hayne, H. (2010). It's not what you know, it's how you know: older siblings facilitate imitation during infancy. *International Journal of Early Years Education, 11*(1), 7–21.
Barrett, M. (1995). *Early lexical development.* In P. Fletcher, & B. MacWhinney (Eds.), *Handbook of Child Language*, (pp. 362 – 392). Oxford: Blackwell.
Barton, M.E. & Tomasello, M. (1991). Joint attention and conversation in mother-infant-sibling triads. *Child Development, 62*(3), 517–529.
Bates, E., Dale, P. & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher, & B. MacWhinney (Eds.), *Handbook of Child Language*, (pp. 96–151). Oxford: Blackwell.
Behrens, H. (2008). Corpora in language acquisition research: history, methods, perspectives. In: Behrens, Heike (Ed.). *Corpora in Language Acquisition Research: Finding Structure in Data,* (pp. xi-xxx). Amsterdam: Benjamin
Ben-Itzchak, E., Zukerman, G. & Zachor, D.A. (2016). Having older sibling is associated with less severe social communication symptoms in young children with autism spectrum disorder. *Journa of Abnormal Child Psychology, 44*, 1613–1620.
Bishop, D.V.M. (1997). Pre- and perinatal hazards and family background in children with specific language impairments: A study of twins. *Brain and Language, 56,* 1–26.
Bishop, D. M., & Bishop, S. J. (1998). `Twin language:' a risk factor for language impairment?. *Journal of Speech, Language & Hearing Research*, *41*(1), 150.
Blackwell, A., Babayigit, S. & Roulstone, S. (2013). *Methodological considerations when using technology for automate vocal analysis (LENA) with young children*. Paper presented at the joint Annual Conference of the British Psychological Society's Developmental and Cognitive Sections in Reading, United Kingdom.

Bland, J.M. & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*, 307–310.
Bland, J.M. & Altman, D.G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical research, 8*, 135–160.
Boersma, Paul & Weenink, David (2014). *Praat: doing phonetics by computer* [Computer program]. Version 5.3.59, retrieved 20th of October 2014 from http://www.praat.org/
Bradley, R.H. & Corwyn, R.F. (2002). Socioeconomic status and child development. *Psychology, 53,* 371–399.
Braid Pharr, A., Bernstein Ratner, N. & Rescorla, L. (2000) Syllable structure development of toddlers with expressive specific language impairment. *Applied Psycholinguistics, 21,* 429 – 449.
Brody, G. (2004). Siblings' direct and indirect contributions to child development. *Current Directions in Psychological Science, 13*(3), 124– 126.
Brown, R. (1973). *A First language: the early stages*. Cambridge: Harvard University Press.
Bruner, J. (1983). *Child's talk: learning to use language*. London: Oxford University Press.
Butler, S., McMahon, C. & Ungerer, J.A. (2003). Maternal speech style with prelinguistic twin infants. *Infant and Child Development, 12,* 129–143.
Canault, M., Le Normand, M., Foudil, S., Loundon, N., & Thai-Van, H. (2015). Reliability of the language Environment analysis system (LENA) in European French. *Behavior Research Methods,*1-16. doi:10.3758/s13428-015-0634-8
Candelori, C., Trumello, C., Babore, A., Keren, M. & Romanelli, R. (2015). The experience of premature birth for fathers: the application of the Clinical Interview for Parents of High-Risk Infants (CLIP) to an Italian sample. *Frontiers in Psychology, 6*, 1–9.
Caskey, M., Stephens, B., Tucker, R. & Vohr, B. (2011). Importance of parent talk on the development of preterm infant vocalizations. *Pediatrics, 128*(5). doi:10.1542/peds.2011-0609
C. Barras, E. Geoffrois, Z. Wu, and M. Liberman. (2000). *Speech communication, 33*, 5–22.
Che, E.S., Alacorn, M.F., Yannaco F.D. & Brooks, P.J. (2016). *Maternal overlap predicts language outcomes for typical and late-talking children*. Conference paper, to appear in Proceedings of the Boston University Conference on Language Acquisition, BULCD 40. [Retrieved 16.6.2016] Available from: https://www.researchgate.net/publication/284563990_Maternal_Overlap_Predicts_Language_Outcomes_for_Typical_and_Late-Talking_Children
Chin, I., Vosoughi, S., Goodwin, M., Roy, D. and Naigles, L. (2013). *Dense data collection through the speechome recorder better reveals developmental trajectories*. Paper presented in the Extended Abstract of the International Meeting for Autism Research (IMFAR) 2013. San Sebastián, Spain.
Christakis DA, Gilkerson J, Richards J.A., Zimmerman, F.J., Garrison, M.M., Xu, D., Gray, S. & Yapanel, U. (2009). Audible television and decreased adult words, infant vocalizations, and conversational turns: A population-based study. *Archives of Pediatrics & Adolescent Medicine, 163*(6), 554-558. doi:10.1001/archpediatrics.2009.61
Conway, D., Lytton, H. & Pysh, F. (1980). Twin-singleton language differences. *Canadian Journal of Behavioral Science, 12*(3), 264–271.
Cristia, A. (2015). *Language Acquisition: From the lab to the wild.* Paper presented in the Workshop on Infant Language Development. Stockholm 10–12.6.2015.
Dale, P.S. & Penfield, M. (2011). *Adaptations of the MacArthur-Bates CDI in to non-US English languages*. Retrieved 3.2.2016 from http://mb-cdi.stanford.edu/documents/AdaptationsSurvey7-5-11Web.pdf

Davis, B.L., McNeilage, P.F., Matyear, C.L. & Powell, J.K. (2000). Prosodic correlates of stress in babbling: An acoustical study. *Child Development, 71*(5), 1258–1270.

Davis, E.A. (1937). *The development of linguistic skills in twins, singletons with siblings and only children from age five to ten years*. Monograph 14. Institute of Child Welfare, University of Minnesota.

Day, E. J. (1932a). The development of language in twins. A comparison of twins and single children. *Child Development 3,* 179–199.

Day, E. J. (1932b). The development of twins: their resemblances and differences. *Child Development 3*, 298–316.

De Bruin, C.D.E., van der Lugt, N.M., Visser, R., Oostdjik, W., van Zwet, E.W., te Pas, A.B. & Lopriore, E. (2015). Dysglycaemia in small-for-gestational-age neonates: a matched case-control study in monochoriotic twins. *The Journal of Maternal-Fetal & Neonatal Medicine*, 1–6.

De Schuymer, L., De Groote, I., Beyers, W., Striano, T. & Roeyers, H. (2011). Preverbal skills as mediators for language outcome in preterm and full term children. *Early Human Development, 87*(4), 265–272.

Dollaghan, C.A., Campbell, T.F., Paradise, J.L., Feldman, H.M., Janosky, J.E., Pitcairn D.N. & Kurs-Lasky, M. (1999). Maternal education and measures of early speech and language. *Journal of Speech, Language, and Hearing Research, 42*, 1432 – 1443.

Duff, F.J., Nation, K., Plunkett, K. & Bishop, D.V.M. (2015). Early prediction of language and literacy problems: is 18 months too early? *PeerJ* 3:e1098. DOI 10.7717/peerj.1098

Duffy, J.R., Watts, J. & Duffy, R.J. (1981). Path analysis: A strategy for investigating multivariate causal relationships in communication disorders. *Journal of Speech and Hearing Research, 24,* 474–490.

Dunn, J. (2006). Siblings and socialization. In J. Grusec & P. Hastings (Eds.). *Handbook of Socialization: Theory and Research*, (pp. 309–327). New York: Guildford Press.

Dunn, J. & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having and older sibling. *Child Development, 60*, 399–410.

Dunst, C.J., Valentine, A. Raab, M. & Hamby, D.W. (2013). Everyday activities as sources of language learning opportunities for infants, toddlers, and preschoolers. *Everyday Child Language Learning Reports, 6*, 1–7.

Dykstra, J. R., Sabatos-DeVito, M. G., Irvin, D. W., Boyd, B. A., Hume, K. A., & Odom, S. L. (2012). Using the language environment analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism, 17*(5), 582-594. doi:10.1177/1362361312446206

Elo, H. & Korpijaakko-Huuhka, A-M (2011). *The phonetic development of a dizygotic twin pair acquiring Finnish*. Poster presented in The 2nd Nordic Conference of Clinical Linguistics, Gran 5–7th of May, 2011.

Ely, R. & Gleason, J.B. (1995). Socialization across contexts. In P. Fletcher, & B. MacWhinney (eds.), *Handbook of Child Language*, 251 - 270. Oxford: Blackwell.

Eusebi, P. (2013), Diagnostic Accuracy Measures. *Cerebrovascular Diseases, 36,* 267–272.

Fagan, M. K. (2009). Mean length of utterance before words and grammar: Longitudinal trends and developmental implications of infant vocalizations. *Journal of Child Language, 36*(3), 495-527.

Falkus, G., Tilley, C., Thomas, C., Hockey, H., Kennedy, A., Arnold, T., Thorburn, B., Jones, K., Bhavika, P., Pimenta, C., Shah, R., Tweedie, F., O'Brien, F., Earney, R. & Pring, T. (2015). Assessing the effectiveness of parent-child interaction therapy with language delayed children: A clinical investigation. *Child Language Teaching and Therapy,* 1–11.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Development, 71*(2), 310-322. doi:10.1111/1467-8624.00146

Felson, J. (2014). What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Social Science Research, 43*, 184–99.

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Reply: measuring variability in early child language: Don't shoot the messenger. *Child Development, 71*(2), 323-328. doi:10.1111/1467-8624.00147

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J. & Pethick, S.J. (1994). Variability on early communicative development. *Monograph of the Society for Research in Child Development, 59*(5), 1–173.

Fernald, A., Taeschner, T., Papousek, M., de Boysson-Bardies, B. & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477–501.

Fletcher, P. (2014). Data and beyond. *Journal of Child Language, 41*(S1), 18–25.

Ford, M., Baer, C.T., Xu, D., Yapanel, U. & Gray, S. (2008). *The LENA™ Language Environment Analysis System: Audio Specifications of the DLP-0121.* (LENA Technical Report LTR-03-02). LENA Foundation, Boulder: Colorado.

Forget-Dubois, N., Dionne, G., Lemelin, J., Perusse, D., Tremblay, R. & Boivin, M. (2009). Early child language mediates the relation between home environment and school readiness. *Child Development, 80*(3), 736–749.

Foster-Cohen, S., Edgin, J.O., Champion, P.R. & Woodward, L.J. (2007). Early delayed language development in very preterm infants: Evidence from the Mac-Arturh-Bates CDI. *Journal of Child Language, 34*, 655–675.

Frank, M.C. (2012). *Computational models of early language acquisition*. Unpublished manuscript. Read 4.3.2016 online from http://langcog.stanford.edu/papers/F-underreview-b.pdf

Frank, M.C., Braginsky, M., Yurovsky, D. & Marchman, V.A. (In revision). Wordbank: An open repository for developmental vocabulary data. Retrieved from http://langcog.stanford.edu/papers_new/frank-2015-underreview.pdf

Frank, M.C., Sugarman, E., Horowitz, A.C., Lewis, M.L. & Yurovsky, D. (2016). Using tablets to collect data from young children. *Journal of Cognition and Develolpment, 17*(1), 1–17.

Ganek, H. & Eriks-Brophy, A. (203*). Questions in using LENA with Vietnamese and Canadian children with hearing loss.* Paper Presented at the LENA International Conference 2013, Denver, CO.

Garitte, C., Almodovar, J-P., Benjamin, E. & Canhao, C. (2002). Speech in same- and different-sex twins 4 and 5 years old. *Twin Research, 5*(6), 538–543.

Gilkerson, J., Coulter, K. & Richards, J.A. (2008). *Transcriptional analyses of the LENA natural language corpus*. (LENA Technical Report LTR-06-02). LENA Foundation, Boulder: Colorado.

Gilkerson, J. & Richards, J. A. (2008). *The LENA™ developmental snapshot.* LENA (Technical Report LTR-07-02). LENA Foundation, Bolder: Colorado.

Gilkerson, J., Zhang, Y., Xu, D., Richards, J.A., Xu, X., Jiang, F., Harnsberger, J. & Topping, K. (2014). Evaluating LENA system performance for Chinese: A pilot study in Shanghai. *Journal of Speech, Language and Hearing research.*

Goldstein, M.H., King, A.P. & West, M. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences of United States, 100*(13), 8030 – 8035.

Golinkoff, R.M., Can, D.D., Soderstrom, M. & Hirsh-Pasek, K. (2015). (Baby)talk to me: the social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Sciences, 24*(5), 339–344.
Goodman, J., Dale, P. & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515–531.
Goodwyn, S.W., Acredolo, L.P. & Brown, C.A. (2000). Impact of symbolic gesturing on early language development. *Journal of Nonverbal Behavior, 24*(2), 81–103.
Goshen-Gottstein, E. (1981). Differential socialization of opposite-sexed twins, triplets, and quadruplets. *Child Development, 52*(4), 1255–1264.
Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly, 32*(2), 83-92.
Gros-Louis, J., West, M.J. & King, A.P. (2014). Maternal responsiveness and the development in directed vocalizing in social interactions. *Infancy, 19*(4), 385–404.
Guimarães, I. & Gouveia, S. (2007). *Do Portugese women have a lower pitch than other female subjects?* Paper presented at the 27th World Congress of IALP, 5.-9.8.2007 Copenhagen, Denmark.
Guo, L.Y. & Eisenberg, S. (2015). Sample length affects the reliability of language sample measures in 3-year olds: Evidence from parent-elicited conversational samples. *Language, Speech, and Hearing Services in Schools,* 46, 141–153.
Haapsamo, H., Kuusikko-Gauffin, S., Ebeling, H., Larinen, K., Penninkilampi-Kerola, V., Soini, H. & Moilanen, I. (2013). Communication development and characteristcs of influencing factors: a follow-up study from 8 to 36 months. *Early Child Development and Care, 183*(2), 321–334.
Haber, M. & Barnhart, H.X. (2006). Coefficients of agreement for fixed observers. *Statistical Methods in Medical Research, 15*, 255–271.
Hart, B. & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes Publishing.
Haworth, C. M. A., Kovas, Y., Harlaar, N., Hayiou-Thomas, M., Petrill, S. A., Dale, P. S., & Plomin, R. (2009). Generalist genes and learning disabilities: A multivariate genetic analysis of low performance in reading, mathematics, language and general cognitive ability in a sample of 8000 12-year-old twins. *Journal of Child Psychology and Psychiatry, 50*(10), 1318-1325.
Hayes, J.R. & Hatch, J.A. (1999). Measuring reliability: Correlation versus percentage of agreement. *Written Communication, 16*(3), 354–367.
Hayiou-Thomas, M., Dale, P. & Plomin, R. (2012). The etiology of variation in language skills changes with development: a longitudinal twin study of language from 2 to 12 years. *Developmental Science, 15*(2), 233–249.
Hedenbro, M. & Rydelius, P-A. (2014). Early interaction between infants and their parents predicts social competence at the age of four. *Acta Paediatrica, 103*, 268 – 274.
Heilmann, J., Ellis Weismer, S., Evans, J. & Hollar, C. (2005). Utility of the MacArthur-Bates communicative development inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology, 14,* 40–51.
Heilmann, J., Miller, J. & Nockerts, N. (2010b). Using language sample databases. *Language, Speech, and Hearing Services in Schools, 41,* 84 –91.
Heilmann, J., Nockerts, N. & Miller, J. (2010a) Language sampling: does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41,* 393 – 404.

Helasvuo, M-L. (2008). Aspects of the structure of Finnish. In A. Klippi & K. Launonen (Eds.), *Research in Logopedics: Speech and Language Therapy in Finland*, 9 – 18. Clevedon: Multilingual matters.
Hills, T. (2013). The company that words keep: comparing the statistical structure of child- versus adult-directed speech. *Journal of Child Language, 40*(3), 586 – 604.
Hinrichs, E. & Krauwer, S. (2014). *The CLARIN Research Infrastructure: Resources and Tools for eHumanities Scholars*. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014*)*, 1525–1531.
Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55–88.
Holditch-Davis, D., Roberts, D. & Sandelowski, M. (1999). Early parental interactions with and perceptions of multiple birth infants. *Journal of Advanced Nursing, 30*(1), 200–210.
Howe, N., Ross, H.S. & Recchia, H. (2011). Sibling relations in early and middle childhood. In P.K Smith & C.H. Hart (Eds.), *The Wiley-Blackwell Handbook of Childhood Social Development*, (pp. 356 – 372). New York, NY: Wiley.
Hua, Z. & Dodd, B. (2000). Phonological systems of a set of Putonghua-speaking twins. *International Journal of Language and Communication Disorders, 35*(4), 487–506.
Hurtado, N., Marchman, V.A. & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*(6), F31–F39.
Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991).Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236–248.
Ingram, D. (1989). *First language acquisition: method, description and explanation*. New York: Cambridge University Press.
Jackson, C.W., Callender, M., Diehm, E. & Meissier, J. (2013). *Spanish-English speaking pre-schoolers: An exploratory study examining LENA data*. Paper Presented at the LENA International Conference 2013, Denver, CO.
Jahn-Samilo, J., Goodman, J., Bates, E. & Sweet, M. (2000). *Vocabulary learning in children from 8 to 30 months of age: a comparison of parental reports and laboratory measures.* (Technical report CND-0006). University of California, San Diego.
Jakobson, R. (1968). *Child language, aphasia and phonological universals*. Paris: Mouton.
Jansson-Verkasalo, E., Valkama, M., Vainionpää, L., Pääkkö, E., Ilkko, E. & Lehtihalmes, M. (2004).Language development in very low birth weight preterm children: A follow-up study. *Folia Phoniatrica et Logopaedica, 56*, 108–119.
Kalashnikova, M. (2015). *Early speech production*. Special session presentation in the Workshop on Infant Language Development. Stockholm 10–12.6.2015.
Kaprio, J. (2006). Twin studies in Finland 2006. *Journal of Twin Research and Human Genetics, 9*(6), 772-777.
Karikoski, R. (2011) Istukan histopatologisen tutkimuksen merkitys. *Duodecim*, *127*(5), 464 – 472.
Keinänen, L. (2010). *Identtisten ja epäidenttisten kaksosten puhe: Monotsygoottisten ja ditsygoottisten kaksosten puheen vertailua kesto-, nopeus-, tauko- ja taajuusmittauksin.* (Academic Dissertation). Retrieved from HELDA-database (http://urn.fi/URN:ISBN:978-952-10-6174-5).
Kent, R., D., Osberger, M.J., Netsell, R. & Goldschmidt Hustedde, C. (1987). Phonetic development in identical twins differing in auditory function. *Journal of Speech and Hearing Disorders, 52*, 64–75.
Kerby, D.S. (2014). The simple difference formula: an approach to teaching non-parametric correlation. *Innovative Teaching, 3*, 1–9.

Keren-Portnoy, T., Majorano, M. & Vihman, M. (2009). From phonetics to phonology: The emergence of first words in Italian. *Journal of Child Language, 36,* 235-267.
Kern, S. & Gayrad, F. (2007). Influence of preterm birth on early lexical and grammatical acquisition. *First Language, 27*(2), 159–173.
Kobayashi, Y., Hayakawa, K., Hattori, R., Ito, M., Kato, K., Hayashi, C. & Mikami, H. (2006). Linguistic features of Japanese twins at 3 or 4 years of age evaluated by Illinois test of psycholinguistic abilities. *Twin Research and Human Genetics, 9*(2), 272-278.
Koopmans-van Beinum, F.J., Clement, C.J. & van den Dikkenberg-Pot, I. (2001). Babbling and the lack of speech perception: a matter of coordination? *Developmental science, 4* (1), 61–70.
Korja, R., Latva, R. & Lehtonen, L. (2011). The effects of preterm birth on mother-infant interaction and attachement during the infant's first two years. *Acta Obstetricia et Gynecologica Scandinavica, 91*, 164–173.
Korpilahti, P., Koljonen, A. & Jansson-Verkasalo, E. (2016). Identification of biological and environmental risk factors for language delay: The Let's Talk STEPS Study. *Infant Behavior and Development, 42*, 27–35.
Kovas, Y., Hayiou-Thomas, M. E., Oliver, B., Dale, P. S., Bishop, D. V. & Plomin, R., (2005). Genetic influences in different aspects of language development: the etiology of language skills in 4.5 year-old twins. *Child Development, 76*(3), 632─651.
Kramer, L. & Conger, K.J. (2009). What we learn from our sisters and brothers: for better or for worse. In L. Kramer & K.J. Conger (Eds.), *Siblings as agents of socialization. New Directions for Child and Adolescent Development*, (1 – 12). San Francisco: Jossey-Bass.
Kuhl, P. (2004). Early language acquisition: Cracking the speech code. *Nature reviews Neuroscience, 5,* 831–843.
Kulick, D. & Schieffelin, B.B. (2004). Language socialization. In A. Duranti (Ed.), *A companion to Linguistic Anthropology*, (pp. 349–368). Oxford, UK: Blackwell.
Kunnari, S., Yliherva, A., Paavola, L. & Peltoniemi, O.M. (2012). Expressive language skills in Finnish two-year-old-extremely- and very-low-birth-weight preterm children. *Folia Phoniatrica et Logopaedica, 64*(1), 5–11.
Laakso, M-L., Eklund, K. & Poikkeus, A-M. (2011). *Esikko – lapsen esikielellisen kommunikaation ja kielen ensikartoitus.* Jyväskylä: Niilo Mäki Institute.
Latvala, A., Rose, R. J., Pulkkinen, L., Dick, D. M. and Kaprio, J. (2014). Childhood verbal development and drinking behaviors from adolescence to young adulthood: a discordant twin-pair analysis. *Alcoholism: Clinical and Experimental Research*, *38*(2), 457–465. doi: 10.1111/acer.12254
Launonen, K. (1987). *Kaksosten kielelliset taidot: viisivuotiaiden riskikaksosten puheen ja kielen tutkimus*. (Unpublished master's thesis). University of Helsinki.
Laurie, R. (2009). Overcoming challenges and succeeding in PISA Science 2006. In R.W. Bybee & B. McCrae (Eds.), *PISA Science 2006: Implications for Science Teachers and Teaching,* (pp 91–100). Arlington: NSTA Press.
Law, J. & Roy, P. (2008). Parental report on infant language skills: a review of the development and application of the communicative development inventories. *Child and Adolescent Mental Health, 13*(4), 198–206.
Lee, S., Davis, B. L. & MacNeilage, P.F. (2009). Universal production patterns and ambient language influences in babbling: A cross-linguistic study of Korean- and English-learning Infants. *Journal of Child Language*, *35,* 591–617.
Leffel, K. & Suskind, D. (2013). Parent-directed approaches to enrich the early language environments of children living in poverty. *Seminars in Speech and Language, 34*, 267–278.

Lehtinen, T. (2014). *Äidin kaksosilleen suuntaama puhe arkisissa vuorovaikutustilanteissa*. (Master's Thesis) Retrieved from Tampub-database (http://urn.fi/URN:NBN:fi:uta-201407102007).
LENA Foundation (2008). *Clinical applications of the breakthrough LENA™ system: monitoring language in the natural environment of the child*. LENA Foundation. Boulder, Colorado. Retrieved from https://www.lenafoundation.org/wp-content/uploads/2014/10/LTR-09-1_Clinical_Applications.pdf
LENA Foundation (2014). *LENA Pro*. [Introductional information about LENA Pro software] Retrieved 9th of January2015 form http://www.lenafoundation.org/lena-pro/.
Leve, L., Harold, G.,T., Ge, X., Neiderhiser, J.M. & Patterson, G. (2010). Refining intervention-targets in family-based research: lessons from quantitative behavioral genetics. *Perspectives on Psychological Science, 5*(5), 516-526. doi:10.1177/1745691610383506
Lieven, E. & Behrens, H. (2012). Dense sampling. In E. Hoff (Ed.), *Research Methods in Child Language. A Practical Guide*, (pp. 226–239). Oxford: Blackwell Publishing.
Li, L., Vikani, A., Harris G.C. & Ling, F.R. (2014). Feasibility study to quantify the auditory and social environment of older adults using a digital language processor. *Otology & Neurotology, 35*(8), 1301 – 1305.
Lung, F., Shu, B., Chiang, T., & Lin, S. (2009). Twin-singleton influence on infant development: A national birth cohort study. *Child: Care, Health and Development, 35*(3), 409-418. doi:10.1111/j.1365-2214.2009.00963.x
Luniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, E. … Kunnari, S. (2015). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods, (August),* 1–24.
Luoto, R., Kinnunen, T., Koponen, P., Kaaja, R., Männistö, S. & Vartiainen, E. (2004) Naisten lisääntymisterveys Suomessa – Tuloksia Finrisk 2002-tutkimuksesta. *Suomen Lääkärilehti, 59*(9), 885 – 890.
Luria, A. R. & Yudovitch, F. I. (1959*). Speech and the development of mental processes in the child.* London: Staples Press.
Luyster, R., Qiu, S., Lopez, K. & Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur-Bates Communicative Development Inventory. *Journal of Speech, Language, and Hearing Research, 50*(3), 667–681.
Lytton, H. (1980). Parent-child interaction: *The Socialization Process Observed in Twin and Singleton Families*. New York: Plenum Press.
Lytton, H., Conway, D. & Sauve, R. (1977). The impact of twinship on parent-child interaction. *Journal of Personality and Social Psychology, 35*(2), 97–107.
Lytton, H. & Gallagher, L. (2008). Parenting twins and the genetics of parenting. In M.H. Bornstein (Ed.*). Handbook of parenting: Volume I: Children and parenting*, 227 - 254. Mahwah, New Jersey: Lawrence Erlbaum.
Lyytinen, P. (1999). *Varhaisen kommunikaation ja kielen kehityksen arviointimenetelmä (MCDI).* Jyväskylä: Yliopistopaino. Niilo Mäki Institute.
Lyytinen, P. (2003). *Morfologiatesti*. Jyväskylä: Kopijyvä Oy. Niilo Mäki Insitute.
Lyytinen, P., Ahonen, T., Eklund, K. ja Lyytinen, H. (2000) *Ääntelyn ja motoriikan kehityksen seurantamenetelmä*. Jyväskylä: Yliopistopaino. Niilo Mäki Institute
Lyytinen, P. & Lyytinen, H. (2004). Growth and predicative relations of vocabulary and inflectional morphology in children with and without familial risk for dyslexia. *Applied Psycholinguistics, 25*, 397–411.
Lyytinen, P., Poikkeus, A-M., Laakso, M-L., Eklund, K. & Lyytinen, H. (2001). Language development and symbolic play in children with and without familial risk for dyslexia. *Journal of speech, Language, and Hearing Research, 44*(4), 873 – 885.

Lyytinen, P., Poikkeus, A-M., Leiwo, M., Ahonen, T. & Lyytinen, H. (1996). Parents as informants of their child's vocal and early language development. *Early Child Development and Care, 126*, 15 – 25.

Löfkvist, U. (2016). *Words make a difference – evaluation of listening and spoken language environment (LENA) in relation to vocabulary development in 0-4 years-old children with hearing impairment and in comparison to age-matched children with normal hearing.* Talk presented in 1st Nordic LENA meeting, Stockholm, 08.04.2016.

MacCoby, E. (2006). Historical overview of socialization research and theory. In Grusec, Joan E., and Hastings, Paul D., (eds.) *Handbook of Socialization: Theory and* Research, (pp. 13 – 41). New York: Guilford Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk.* Lawrence Erlbaum, Mahwah, NJ.

MacWhinney, B. (2014). What we have learned. *Journal of Child Language, 41* (S1), 124–131.

Mann, J. (1992). Nurturance or negligence: maternal psychology and behavioral preference among preterm twins. In J.H. Barkov, L. Cosmides & J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture,* (367–390). New York: Oxford University Press.

Mayr, R., Price, S. & Mennen, I. (2012). First language attrition in the speech of Dutch-English bilinguals: the case of monozygotic twin sisters. *Bilingualism: Language and Cognition, 15*(4), 687–700.

McCartney, K. (1984). Effect of quality of day care environment on children's language development. *Developmental Psychology, 20*(2), 244–260.

McCathren, R.B., Yoder, P.J. & Warren, S. (1999). The relationship between prelinguistic vocalization and later expressive vocabulary in young children with developmental delay. *Journal of Speech, Language and Hearing Research, 42*, 915–924.

McEvoy, S. & Dodd, B. (1992). The communication abilities of 2- to 4-year-old twins. *European Journal of Communication Disorderds, 27*(1), 73–87.

McGillion, M., Herbert, J.S., Pine, J., Vihman, M., dePaolis, R., Keren-Portnoy, T. & Mathews, D. (In press). What paves the way to conventional language? The predictive value of babble, pointing and SES. *Child Development, e-publication ahead of print.* Available from: http://eprints.whiterose.ac.uk/id/eprint/92982

McGowan, J.E., Alderdice, F.A., Holmes, V.A. & Johnston, L. (2011). Early childhood development of late-preterm infants: a systematic review. *Pediatrics, 127*(6), 1111–1124.

McLeod, S. & Bleile, K. (2004). The ICF: a framework for setting goals for children with speech impairment. *Child Language Teaching and Therapy, 20*(3), 199–219.

McMahon, S., Stassi, K. & Dodd, B. (1998). The relationship between multiple birth children's early phonological skills and later literacy. *Language, Speech, and Hearing Services in Schools, 29,* 11 – 23.

Miceli, P.J., Goeke-Morey, M.C., Whitman, T.L., Sipes Kolberg, K., Miller-Loncar, C. & White, R.D. (2000). Brief report: birth status, medical complications, and social environmnet: individual differences in development of preterm, very low birth weight infants. *Journal of Pediatric Psychology, 25*(5), 353–358.

Minde, K., Corter, C., Goldberg, S. & Jeffers, D. (1990). Maternal preference between premature twins up to age four. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*(3), 367–374.

Mittler, P. (1970). Biological and social aspects of language development in twins. *Developmental Medicine and Child Neurology, 12*, 741 – 757.

Mogford, K. (1993). Language development in twins. In D. Bishop & K. Mogford (Eds.), *Language development in exceptional circumstances,* (pp. 80—95). Hove: Lawrence Erlbaum.

Mogford-Bevan, K. (1999). Twins and language development. In A.C. Sandbank (Ed.), *Twin & Triplet Psychology: a professional guide to working with multiples*, (pp. 36–60). London: Routledge.
Mogford-Bevan, K. (2000). Developmental language impairments with complex origins: Learning from twins and multiple children. *Folia Phoniatrica et Logopaedica, 52,* 74–82.
Molemans, I., van den Berg, R., van Severen, L. & Gillis, S. (2012). How to measure the onset of babbling reliably. *Journal of Child Language, 39,* 523–552.
Morris, S.R. (2010). Clinical application of the mean babbling level and syllable structure level. *Language, Speech, and Hearing Services in Schools, 41*, 223–230.
Muller-Nix, C., Forcada-Guuex, M., Pierrehumbert, B., Jaunin, L., Borghini, A. & Ansermet, F. (2004). Prematurity, maternal stress and mother-child interactions. *Early Human Development, 79*, 145–158.
Nan, C., Piek, J., Werner, C., Mellers, D., Krone, R.E., Barret, T. & Zeekers, M.P. (2013). Trajectories and predictors of developmental skills in healthy twins up to 24 months of age. *Infant Behavior and Development, 36*(4), 670–678.
Odean, R., Nazareth, A. & Pruden, S.M. (2015). Novel methodology to examine cognitive and experimental factors in language development: combining eye tracking and LENA technology. *Frontiers in Psychology, 6*, 1–12.
Oetting, J. B., Hartfield, L. R., & Pruitt, S. L. (2009). Exploring LENA as a tool for researchers and clinicians. *ASHA Leader, 14*(6), 20-22.
Official Statistics of Finland (OSF): Educational structure of population [e-publication]. ISSN=2242-2919. Helsinki: Statistics Finland [referred: 9.3.2016]. Access method: http://www.stat.fi/til/vkour/index_en.html
Okeh, U.M. & Okoro, C.N. (2012). Evaluating measures of indicators of diagnostic test performance: fundamental meanings and formulars. *Journal of Biometrics and Biostatistics, 3*(1), 132–142.
Oller, D.K. & Eilers, R.E. (1988). The role of audition in infant babbling. *Child Development, 59,* 441–449.
Oller, D.K., Eilers, R.E., Neal, A.R. & Cobo-Lewis, A.B. (1998). Late onset of canonical babbling: A possible early marker for abnormal development. *American Journal of Mental Retardation, 103*(3), 249–263.
Oller, D. K., Eilers, R.E., Neal, A.R, & Schwartz, H.K. (1999). Precursors to speech in infancy: The prediction of speech and language disorders. *Journal of Communication Disorders, 32,* 223–245.
Oller, D.K., Levine, S.L., Cobo-Lewis, A.B., Eilers, R.E. & Zurer Pearson, B. (1998). Vocal precursors to linguistic communication. How babbling is connected to meaningful speech. In R. Paul (Ed.). *Exploring the Speech-Language Connection*, (pp. 1–23). Baltimore, MD: Paul H. Brookes.
Oller, D.K., Niyogi P., Gray, S., Richards, J.A., Gilkerson, J., Xu, D., Yapanel, U. Warren, S.F. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences of the United States of America, 107, 30,* 13354–13359.
Oshima-Takane, Y., Goodz, E. & Derevensky, J.L. (1996). Birth order effects on early language development: Do secondborn children learn from overheard speech. *Child Development, 67*, 621–63.
Oshima-Takane, Y. & Robbins, M. (2003). Linguistic environment of secondborn children. *First Language, 23* (1), 21–40.

Pae, S. (2013). *An investigation of the language environment of young Korean children and the impact of feedback on language development*. Paper Presented at the LENA International Conference 2013, Denver, CO.
Partanen, L., Olsén, P., Mäkikallio, K., Korkalainen, N., Heikkinen, H., Heikkinen, M. & Yliherva, A. (2016). Communication profile of primary school-aged children with foetal growth restriction. *Child Language Teaching and Therapy.* doi:10.1177/0265659016654205
Paul, R., Sgangle Looney, S. & Dahm, P.S. (1991). Communication and socialization skills at ages 2 and 3 in "late-talking" young children. *Journal of Speech and Hearing Research, 34,* 858–865.
Paul, R. & Jennings, R. (1992). Phonological behaviour in toddlers with slow expressive language development. *Journal of Speech, and Hearing Research, 35*, 99–107.
Pertilä, P., Hämäläinen, Mieskolainen, (2013). Passive temporal offset estimation of multichannel recordings of an ad-hoc michrophone array, *IEEE Transactions on Audio, Speech and Language Processing, 21*(11), 2393–2402.
Petäjistö, L. (2016). *Kolmevuotiaiden isosiskojen kielenkäyttö seitsenkuisten kaksossisarusten kieliympäristönä. Kaksi tapaustutkimusta*. (Master's Thesis). Retrieved from Tampub-database
http://tampub.uta.fi/bitstream/handle/10024/98597/gradu07285.pdf?sequence=1
Pickstone, C., Goldbart, J., Marshall, J., Rees, A. & Roulstone, S. (2009). A systematic review of environmental interventions to improve child language outcomes for children with or at risk of primary language impairment. *Journal of Research in Special Educational Needs, 9*(2), 66–79.
Plomin, R. (2011). Commentary: Why are children in the same family so different? Non-shared environment three decades later. International *Journal of Epidemiology, 40,* 582–592.
Plomin, R. & Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral and Brain Sciences, 10*(01), 1–16.
Purho, J., Nuutila, M. & Heikinheimo, O.(2008). Kaksosraskaudet. *Lääketieteellinen Aikakauskirja Duodecim, 124*(10), 1111–1119.
Radder, H. (2003). *The philosophy of scientific experimentation*. Pittsburgh: University Press of Pittsburgh.
Raddder, H. (2009). The philosophy of scientific experimentation: A review. *Automated Experimentation, 1,* 2. doi:10.1186/1759-4499-1-2
Ramirez-Esparza, N., Garcia-Sierra, A. & Kuhl, P.K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science, 17*(6), 880–891.
Rautakoski, P., Hannus, T., Simberg, S., Sandnabba, K. & Santtila, P. (2012). Genetic and environmental effects on stuttering: A twin study from Finland. *Journal of Fluency Disorders, 37*(3), 202–210.
Rendle-Short, J., Skelt, L. & Bramley, N. (2015). Speaking of twin children: evidence against the 'impoverishment' thesis. *Research on Language and Social Interaction, 48*(1), 79–99.
Rescorla, L. (2005). Age 13 Language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 48,* 459 – 472.
Rice, M. L., Zubrick, S. R., Taylor, C. L., Gayan, J., & Bontempo, D. (2014). Late language emergence in 24-month-old twins: heritable and increased risk for late language emergence in twins. *Journal of Speech, Language, and Hearing Research, 57*, 917-928.
Rosen, C. (2013). We need a Nuremberg code for big data. Slate, retrieved 20.10.2016 from http://www.slate.com/articles/technology/future_tense/2013/06/providence_talks_program_and_the_rise_of_social_engineering_surveillance.html

Rothgänger, H. (2003). Analysis of the sounds of the child in the first year of age and a comparison to the language. *Early Human Development, 75,* 55–69.
Roulstone, S., Law, J., Rush, R., Clegg, J. & Peters, T. (2011). *Investigating the role of language in children's early educational outcomes.* (Research Report DFE-RR134). Retrieved 9.5.2016 from Department of Education, University of the West of England, Bristol. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181549/DFE-RR134.pdf
Rowe, M. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, 25, 185–205.
Rowe. M. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762–1774.
Rowe, D. C., Vesterdal, W. J., & Rodgers, J. L. (1998). Herrnstein's syllogism: genetic and shared environmental influences on IQ, education, and income. *Intelligence, 26*(4), 405-423. doi:http://dx.doi.org/10.1016/S0160-2896(99)00008-2
Roy, D. (2009). New horizons in the study of child language acquisition. *Proceedings of Interspeech 2009*. Brighton, England.
Roy, B.C., Frank, M.C., DeCamp, P., Miller, M. & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences of the United States of America, 112*(41), 12663–12668. doi:10.1073/pnas.1419773112
Roy, B.C., Frank, M.C. & Roy, D. (2012). Relating activity contexts to early word learning in dense longitudinal data. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*. Sapporo, Japan
Roy, D., Patel, R., DeCamp, P., Kubat, R., Fleischman, M., Roy, B., … Gorniak, P. (2006). The Human Speechome åroject. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, July 26–29. Vancouver, Canada.
Rutter, M., Thorpe, K., Greenwood, R., Nothstone, K. & Golding, J. (2003). Twins as a natural experiment to study the causes of mild language delay: I: Design; twin – singleton differences in language and obstetric risks *Journal of Child Psychology and Psychiatry 44*(3), 326—341.
Räisänen, A. (1975). Havaintoja lastenkielestä. *Virittäjä*, 251–266.
Savic, S. (1980). *How twins learn to talk: a study of the speech development of twins*. London: Academic Press.
Saaristo-Helin, K. & Savinainen-Makkonen, T. (2008). Lasten äänteellinen kehitys: Tutkimusmetodologiaa. *Puhe ja kieli, 4*, 1–32.
Saaristo-Helin, K., Kunnari, S. & Savinainen-Makkonen, T. (2011). Phonological development in children learning Finnish: A review. *First Language, 31,* 342–363.
Sachse, S & Von Suchodoletz, W. (2008). Early identification of language delay by direct language assessment or parent report? *Journal of Developmental & Behavioral Pediatrics, 29*, 34–41.
Sacks, C., Shay, S., Repplinger, L., Leffel, K.R., Sapolich, S.G., Suskind, E., Tannenbaum, S. & Suskind, D. (2014). Pilot testing of a parent-directed intervention (Project ASPIRE) for underserved children who are deaf or hard of hearing. *Child Language Teaching and Therapy, 30*(1), 91–102.
Sameroff, A.J. & Fiese, B.H. (2000). Transactional Regulation: The developmental ecology of early interventions. In J.P. Shonkoff, S.J. Meisels (Eds.), *Handbook of Early Childhood Intervention, (*pp. 361- 385). Edinburgh: Cambridge University Press.
Sarkadi, A., Kristiansson, R., Oberklaid, F. & Bremberg, S. (2008). Fathers' involvement and children's developmental outcomes: a systematic review of longitudinal studies. *Acta Paediatrica, 97*, 153–158.

Savic, S. (1980). *How twins learn to talk: a study of the speech development of twins*. London: Academic Press.

Savinainen-Makkonen, T. (2000). Learning to produce three-syllable words. In Perking & Howard (Eds.), *New Directions in Language Development and Disorders,* (pp. 223–231). Plenum Publishers: NY.

Savinainen-Makkonen, T. (2013). Geminate template: a model for first Finnish words. In M.M. Vihman & T. Keren-Portnoy (Eds.), *The Emergence of Phonology: Whole-word Approaches and Cross-Linguistic Evidence*, (pp. 362–374). Croydon: Cambridge University Press.

Schermann-Eizirik, L., Hagekull, B., Bohlin, G., Persson, K. & Sedin, G. (1997). Interaction between mothers and infants born at risk during the first six months of corrected age. *Acta Paediatrica, 86*, 864–872.

Schieffelin, B. & Ochs, E. (2008). Language socialization: an historical overview. In P.A. Duff & H. Hornberger (Eds.), *Encyclopedia of Language and Education, 2nd Edition, Volume 8: Language Socialization*, (pp.3–15). New York: Springer Science.

Schneidman, L.A., Arroyo, M.E., Levine, S.C. & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language, 40*(3), 672–686. DOI: 10.1017/S0305000912000141,

Shapiro-Mendoza, C.K. & Lackritz, E.M. (2012). Epidemiology of late and moderate preterm birth. *Seminars in Fetal and Neonatal Medicine, 17*(3), 120–125.

Sharma, D., Shastri, S. & Sharma, P. (2016). Intrauterine growth restriction- part 1. *Journal of Maternal-Fetal and Neonatal Medicine*, 1–39.

Slobin, D. (2014). Before the beginning: the development of tools of the trade. *Journal of Child Language, 41*(S1), 1–17.

Smith, G.C., Fleming, K.M. & White, I.R. (2007). Birth order of twins and risk of perinatal death related to delivery in England, Northern Ireland, and wales, 1994–2003: retrospective cohort study. *BMJ, 334*, 576. doi:10.1136/bmj.39118.483819.55

Smith, G.C.S., Shah, I., White, I.R., Pell, J.P. & Dobbie, R. (2005). Mode of delivery and the risk of delivery-related perinatal death among twins at term: a retrospective cohort study of 8073 births. *International Journal of Obtetrics and Gynaecology, 112*(8), 1139–1144.

Snow, C. (1995) Issues in the study of input: fine-tuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher, & B. MacWhinney (eds.), *Handbook of Child Language*, (pp. 181–193). Oxford: Blackwell.

Snow, C. (2014). Input to interaction to instruction: three key shifts in the history of child language research. *Journal of Child Language, 41*(S1), 117–123.

Soderstrom, M. (2007). Beyond babytalk: re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*, 501–532.

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *Plos One, 8*(11), e80646.

Stafford, L. (1987). Maternal input to twin and singleton children. Implications for language acquisition. *Human Communication Research, 13*(4), 429–462.

Stakes (2009). Synnytykset ja vastasyntyneet vuonna 2008. Retrieved from http://www.stakes.fi/tilastot/tilastotiedotteet/2009/tr22_09.pdf

Stanton-Chapman, T.L., Chapman, D.A., Bainbridge, N.L. & Scott, K.G. (2002). Identification of early risk factors for language impairment. *Research in Developmental Disabilities, 23*, 390–405.

Stewart, E.A. (2000). Towards the social analysis of twinship. *British Journal of Sociology, 51*(4), 719–737.

Stoel-Gammon, C. (1989). Prespeech and early speech development of two late talkers. *First Language, 9,* 207–223.
Stoel-Gammon, C. (1991). Normal and disordered phonology in two-year olds. *Topics in Language Disorders*, 11, 21–32.
Topics in Language Disorders, 11, 21–32.Stolt, S., Haataja, L., Lapinleimu, H. & Lehtonen, L. (2008).Associations between lexicon and grammar at the end of the second year in Finnish children. *Journal of Child Language, 36*(4), 779–806.
Stolt, S., Haataja, L., Lapinleimu, H. & Lehtonen, L. (2009a). The early lexical development and its predictive to language skills at 2 years in very-low-birth-weight children. *Journal of Communication Disorders, 42,* 107–123.
Stolt, S., Haataja, L., Lapinleimu, H. & Lehtonen, L. (2009b). Associations between lexicon and grammar at the end of the second year in Finnish childrne. *Journal of Child Language, 36*(4), 779–806.
Stolt, S., Lehtonen, L., Haataja, L. & Lapinleimu, H. (2012). Development and predictive value of early vocalizations in very-low-birth-weight children: a longitudinal study. *Clinical Linguistics & Phonetics, 26*(5), 414–427.
Stolt, S., Lind, A., Matomäki, J., Haataja, L., Lapinleimu, H. & Lehtonen, L. (2016). Do early development of gestures and receptive and expressive language predict language skills at 5;0 in prematurely born very-low-birth-weight children? *Journal of Communication disorders, 61*, 16–28.
Stormswold, K. (2001). The heritability of language: A review and meta-analysis of twin, adoption and linkage studies. *Language, 77,* 647–723.
Stormswold, K. (2006). Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition, 101*(2), 333-384. DOI: http://dx.doi.org/10.1016/j.cognition.2006.04.007
SVT (Official Statistics of Finland). Perinatal statistics: parturities, deliveries and newborns 2014. National institute for Health and Welfare. Retrieved from http://www.stat.fi/til/sysyvasy/index.html
Tamis-LeMonda, Kuchirko, Y. & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science, 23* (2), 121–126.
Tammen, S.A., Friso, S. & Choi, S-W. (2013). Epigenetics: the link between nature and nurture. *Epigenetics, 34*(4), 753–764.
Teinonen, T. (2009). *Language learning in infancy.* (Academic Dissertation). Retrieved from HELDA-database (http://urn.fi/URN:ISBN:978-952-10-5875-2).
Teinonen, T. & Huotilainen, M. (2012). Implicit segmentation of a stream of syllables based on transitional probabilities: an MEG study. *Journal of Psycholinguistic Research, 41*(1), 71–82.
Thelen, E. (2005). Dynamic Systems Theory and the Complexity of Change. *Psychoanalytic Dialogues, 15*(2), 255–283.
Thorpe, K. (2006). Twin children's language development. *Early Human Development, 82, 6,* 387 – 395.
Thorpe, K., Rutter, M. & Greenwood, R. (2003). Twins as a natural experiment to study the causes of mild language delay: II: Family interactions risk factors. *Journal of Child Psychology and Psychiatry, 44*(3), 342 – 355.
Tiitinen, A. (2011). *Monisikcöisyys*, Lääkärikirja Duodecim. Retrieved from http://www.terveyskirjasto.fi/terveyskirjasto/tk.koti?p_artikkeli=dlk00154&p_haku=monisiki%C3%B6isyys.
Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Cambridge, MA: Harvard University Press.

Tomasello, M. & Mannle, S. (1985). Pragmatics of sibling speech to one-year-olds. *Child Development, 56,* 911–917.

Tomasello, M., Mannle, S. & Kruger, A.C. (1986). Linguistic environment of 1-2 year old twins. *Developmental Psychology, 22,* 169-176.

Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: how much is enough? *Journal of Child Language, 31*, 101–121.

Trainor, L.J. & Desjardins, R.N. (2002). Pitch characteristics of infant-directed speech affect infant's ability to discriminate vowels. *Psychonomic Bulletin & Review, 9(*2), 335–340.

Trias T, Ebeling H, Penninkilampi-Kerola V, Moilanen I. (2010). Depressive and psychosomatic symptoms in twins with special reference to co-twin dependence. *Twin Research and Human Genetics*, *13* (6), 559-66.

Tremblay-Leveau, H., Leclerc, S. & Nadel, J. (1999). Linguistic skills of 16- and 23 –month-old twins and singletons in a triadic context. *First Language, 19,* 233—254.

Törölä, H., Lehtihalmes, M., Heikkinen, H., Olsèn, P. & Yliherva, A. (2012). Early vocalization of preterm infants with extremely low birth weight (ELBW), part II: from canonical babbling up to the appearance of the first word. *Clinical Linguistics & Phonetics, 26*(4), 345–356.

United Nations Development Programme (UNDP) (2013). Mean years of schooling (of adults) (years). Retrieved from http://hdr.undp.org/en/content/mean-years-schooling-adults-years#footnote

Uotila, J., Vuorela, P., Koistinen, E., Nupponen, I., Nuutila, M., Raudaskoski, T., Tammela, O. & Varesmaa-Korhonen, L. (2011). Ennenaikainen synnytys. Retrieved from http://www.terveysportti.fi.helios.uta.fi/dtk/ltk/koti?p_haku=ennenaikainen%20synnytys

Vainio, K., Haataja, L., Lapinleimu, L., Lehtonen, L., Stolt, S. & PIPARI-tutkimusryhmä (2011). Kahden vuoden iässä mitatun sanaston koon ja koostumuksen yhteys nimeämistaitoihin viiden vuoden iässä. *Puhe ja kieli, 31*(1), 25–42.

VanDam, M., Ambrose, S. E., & Moeller, M. P. (2012). Quantity of parental language in the home environments of hard-of-hearing 2-year-olds. *Journal of Deaf Studies and Deaf Education, 17*(4), 402-420.

VanDan M. & Silbert, N. (2013a). *Fidelity of LENA vocal activity labels*. Paper presented at the LENA Conference 2013, Denver, Colorado.

VanDam, M. & Silbert, N. (2013b). *Precision and error of automatic speech recognition*. Poster presented at the Proceedings of Meetings on Acoustics, vol. 19. ICA 2013. Montreal.

VanDam, M. & Tully, T. (2016). *Quantity of mothers' and fathers' speech to sons and daughters*. Talk presented at the 171st Meeting of the Acoustical Society of America, Salt Lake City, UT, May 25, 2016.

VanDam, M., Warlaumont, A.S, Bergelson, E., Cristia, A., Soderstrom, M., DePalma, P. & MacWhinney, B. (2016). Homebank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language, 37*(02), 128–142.

Van Hulle, C.A., Goldsmith, H.H. & Lemery, K.S. (2004). Genetic, environmental, and gender effects on individual differences in toddler expressive language. *Journal of Speech, Language and Hearing Research, 47,* 904–912.

Vannest, J., Bertram, R., Järvikivi, J. & Niemi, J. (2002). Counterintuitive cross-linguistic differences: more morphological computation in English than in Finnish. *Journal of Psycholinguistic Research, 31*(2), 83–106.

Vanormelingen, L. & Gillis, S. (2016). The influence of socio-economic status on mother's volubility and responsiveness in a monolingual Dutch-speaking sample. *First Language, 36*(2), 140–156.

Van Severen, L., van den Berg, R., Molemans, I. & Gillis, S. (2012). Consonant inventories in the spontaneous speech of young children: A bootstrapping procedure. *Clinical Linguistics & Phonetics, 26*(2), 164–187.
Viera, A.J. & Garrett, J.M. (2005). Understanding interobserver agreement: the Kappa statistic. *Family Medicine, 37*(5), 360–363.
Vihman, M., Ferguson, C.A. & Elbert, M. (1986). Phonological developement from babbling to speech: common tendencies and individual differences. *Applied Psycholinguistics, 7,* 3–40.
Vosoughi, S. & Roy, D. (2012). An automatic child-directed speech detector for the study of child language development. *Proceedings of Interspeech 2012.* Portland, Oregon.
Vygotski, L.S. (1982). *Ajattelu ja kieli.* Espoo: Weilin+Göös.
Wallace, I.F., Berkman, N.D., Watson, L.R., Coyne-Beasley, T., Wood, C.T., Cullen, K. & Lohr, K.N. (2015). Screening for speech and language delay in children 5 Years and younger: A systematic review. *Pediatrics, 136*(2), 448–462.
Wang, Z., Pan, X., Miller, K. F., & Cortina, K. S. (2014). Automatic classification of activities in classroom discourse. *Computers & Education, 78*, 115-123. doi:10.1016/j.compedu.2014.05.010
Warren, S.F. & Gilkerson, J. (2008). *Automatic measurement of the language learning environment of children with ASD*. Paper presented in 39th Conference of Autism Society of America. Orlando, Florida.
Warren, S. F., Gilkerson, J., Richards, J. A., Oller, K. D., Xu, D., Yapanel, U., & Gray, S. (2010). What automated vocal analysis reveals about the vocal production and language learning environment of young children with autism. *Journal of Autism and Developmental Disorders, 40*(5), 555-569.
Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143-2152. doi:10.1177/0956797613488145
Winerman, L. (2004). A second look at twin studies. *Monitor in Psychology, 35*(4), 46.
Woollet, A. (1986). The influence of older siblings on the language environment of young children. *British Journal of Developmental Psychology, 4,* 235–245.
World Health Organization (WHO) (2015). *Preterm birth.* Internet web site. Retrieved from http://www.who.int/mediacentre/factsheets/fs363/en/.
Xu, D. & Yapanel, U. & Gray, S. (2009). *Reliability of the LENA™ Language Environment Analysis System in Young Children's Natural Home Environments*. (LENA Technical Report LTR-05-02). LENA Foundation, Boulder: Colorado.
Xu, D. & Yapanel, U., Gray, S. & Baer, C.T. (2008). *The LENA™ Language Environment Analysis System: The Interpreted Time Segments (ITS) File*. (LENA Technical Report LTR-04-02). LENA Foundation, Boulder: Colorado.
Xu, D., Yapanel, U, Gray, S., Gilkerson, J., Richards, J. & Hansen, J. (2008). *Signal processing for young child speech development*. Paper presented at WOCCI 2008: The 1st Workshop on Child, Computer, and Interaction. Chania, Greece.
Zhang, Y., Xu, X., Jiang, F., Gilkerson, J., Xu, D., Richards, J.A., Harnsberger, J. & Topping, K.J. (2015). Effects of quantitative linguistic feedback to caregivers of young infants: a pilot study in China. *Communication Disorders Quarterly*, 1–9.
Zimmerman, F. J., Gilkerson, J., & Richards, J. A. (2009). Teaching by listening: the importance of adult-child speech conversations to language development. *Pediatrics*, 124–342.
Zlatic, L., MacNeilage, P.F., Matyear, C.L. & Davis, B.L. (1997). Babbling of twins in bilingual environment. *Applied Psycholinguistics, 18*(4), 453–469.

Yinon, Y., Mazkereth, R., Rosentzweig, N., Jarus-Hakak, A., Schiff, E. & Simchen, M.J. (2005). Growth restriction as a determinant of outcome in preterm discordant twins. *Obstetrics and Gynecology, 105*(1), 80–84.

Yliherva, A., Olsén, P., Mäki-Torkko, E., Koiranen, M. & Järvelin, M-R. (2001). Linguistic and motor abilities of low-birthweight children as assessed by parents and teachers at 8 years of age *Acta Paediatrica, 90,* 1440 – 1449.

Yurovsky, D., Doyle, G. & Frank, M.C. (2016). *Linguistic input is tuned to children's developmental level.* Paper presented in the Annual Meeting of the Cognitive Science society, Philadelphia, USA.

# APPENDICES

## Appendix 1. LENA and human agreement, human as a gold standard

*Inter-rater agreement of speaker labelling, R1*

| | | HUMAN | | | | | |
|---|---|---|---|---|---|---|---|
| | | key child | female adult | male adult | other child | not recognized | Total |
| Lena labels | key child | 4649 | 16 | 6 | 427 | 70 | 5168 |
| | | 90.0% | 0.3% | 0.1% | 8.3% | 1.4% | 100.0% |
| | female adult | 87 | 1142 | 34 | 35 | 53 | 1351 |
| | | 6.4% | 84.5% | 2.5% | 2.6% | 3.9% | 100.0% |
| | male adult | 3 | 277 | 511 | 5 | 239 | 1035 |
| | | 0.3% | 26.8% | 49.4% | 0.5% | 23.1% | 100.0% |
| | other child | 113 | 69 | 4 | 2504 | 56 | 2746 |
| | | 4.1% | 2.5% | 0.1% | 91.2% | 2.0% | 100.0% |
| | not recognized | 24 | 6 | 8 | 20 | 6 | 64 |
| | | 37.5% | 9.4% | 12.5% | 31.3% | 9.4% | 100.0% |
| Total | | 4876 | 1510 | 563 | 2991 | 424 | 10364 |
| | | 47.0% | 14.6% | 5.4% | 28.9% | 4.1% | 100.0% |

κ=.775, p=.005

*Inter-rater agreement rates for speaker labelling, R2*

| | | HUMAN | | | | | |
|---|---|---|---|---|---|---|---|
| | | key child | female adult | male adult | other child | not recognized | Total |
| LENA | key child | 3673 | 58 | 9 | 497 | 242 | 4479 |
| | | 82.0% | 1.3% | 0.2% | 11.1% | 5.4% | 100.% |
| | female adult | 161 | 3310 | 49 | 21 | 181 | 3722 |
| | | 4.3% | 88.9% | 1.3% | 0.6% | 4.9% | 100% |
| | male adult | 10 | 90 | 827 | 0 | 154 | 1081 |
| | | 0.9% | 8.3% | 76.5% | 0.0% | 14.2% | 100% |
| | other child | 209 | 209 | 3 | 1431 | 180 | 2032 |
| | | 10.3% | 10.3% | 0.1% | 70.4% | 8.9% | 100% |
| | not recognized | 83 | 67 | 10 | 49 | 0 | 209 |
| | | 39.7% | 32.1% | 4.8% | 23.4% | 0.0% | 100% |
| Total | | 4136 | 3734 | 898 | 1998 | 757 | 11523 |
| | | 35.9% | 32.4% | 7.8% | 17.3% | 6.6% | 100% |

κ=.724, p=.005

*Inter-rater agreements for speech-like vocalizations, R1*

| | | HUMAN | | | | | |
|---|---|---|---|---|---|---|---|
| | | speech-like vocalization | vegetative sound | cry | silence | no label | Total |
| LENA | speech-like vocalization | 676 | 6 | 2 | 34 | 88 | 806 |
| | | 83.9% | .7% | .2% | 4.2% | 10.9% | 100.0% |
| | vegetative sound | 3 | 69 | 0 | 0 | 4 | 76 |
| | | 3.9% | 90.8% | 0.0% | 0.0% | 5.3% | 100.0% |
| | cry | 235 | 26 | 957 | 2 | 115 | 1335 |
| | | 17.6% | 1.9% | 71.7% | .1% | 8.6% | 100.0% |
| | silence | 1 | 2 | 0 | 757 | 31 | 791 |
| | | .1% | .3% | 0.0% | 95.7% | 3.9% | 100.0% |
| | no label | 0 | 0 | 4 | 1 | 2022 | 2027 |
| | | 0.0% | 0.0% | .2% | .0% | 99.8% | 100.0% |
| Total | | 915 | 103 | 963 | 794 | 2260 | 5035 |
| | | 18.2% | 2.0% | 19.1% | 15.8% | 44.9% | 100.0% |

κ=.846, p=.006

*Inter-rater agreements for speech-like vocalizations, R1*

| | | HUMAN | | | | | |
|---|---|---|---|---|---|---|---|
| | | speech-like vocalization | vegetative sound | cry | silence | no label | Total |
| LENA | speech-like vocalization | 798 | 7 | 0 | 0 | 23 | 828 |
| | | 96.4% | .8% | 0.0% | 0.0% | 2.8% | 100.0% |
| | vegetative sound | 8 | 131 | 0 | 0 | 16 | 155 |
| | | 5.2% | 84.5% | 0.0% | 0.0% | 10.3% | 100.0% |
| | cry | 80 | 33 | 621 | 0 | 37 | 771 |
| | | 10.4% | 4.3% | 80.5% | 0.0% | 4.8% | 100.0% |
| | silence | 3 | 0 | 0 | 814 | 25 | 842 |
| | | .4% | 0.0% | 0.0% | 96.7% | 3.0% | 100.0% |
| | no label | 77 | 11 | 4 | 0 | 0 | 92 |
| | | 83.7% | 12.0% | 4.3% | 0.0% | 0.0% | 100.0% |
| Total | | 966 | 182 | 625 | 814 | 101 | 2688 |
| | | 35.9% | 6.8% | 23.3% | 30.3% | 3.8% | 100.0% |

κ=.833, p=.008

Appendix 2. Key child daily durations and neonatal health-related background variables in monthly inspection

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pretermity | 6 | full-term | 12 | 2465.93 | 886.22 | 2316.11 | 1560.48 | 45.00 | -0.23 | 0.851 | 0.06 |
| | | preterm | 8 | 2272.89 | 480.71 | 2246.02 | 518.26 | | | | |
| | 7 | full-term | 14 | 2472.50 | 664.24 | 2436.11 | 901.69 | 47.00 | -0.61 | 0.570 | 0.16 |
| | | preterm | 8 | 2070.65 | 564.39 | 1929.13 | 1801.81 | | | | |
| | 8 | full-term | 14 | 2520.09 | 842.70 | 2342.93 | 1238.40 | 47.00 | -0.61 | 0.570 | 0.16 |
| | | preterm | 8 | 2185.66 | 454.08 | 2148.86 | 529.04 | | | | |
| | 9 | full-term | 12 | 2468.92 | 661.45 | 2406.06 | 864.18 | 40.00 | -0.62 | 0.571 | 0.17 |
| | | preterm | 8 | 2637.28 | 433.05 | 2540.05 | 739.08 | | | | |
| | **10** | **full-term** | **13** | **2136.67** | **455.24** | **2225.21** | **759.71** | **12.00** | **-2.90** | **0.003** | **0.77** |
| | | **preterm** | **8** | **2856.38** | **548.44** | **2817.13** | **641.46** | | | | |
| | **11** | **full-term** | **12** | **2260.45** | **352.33** | **2259.60** | **350.91** | **12.00** | **-2.78** | **0.004** | **0.75** |
| | | **preterm** | **8** | **3091.54** | **791.60** | **2835.15** | **811.95** | | | | |
| | 12 | full-term | 10 | 2455.13 | 769.85 | 2439.97 | 1339.77 | 25.00 | -0.98 | 0.364 | 0.29 |
| | | preterm | 7 | 2754.97 | 627.00 | 2826.80 | 884.84 | | | | |
| birth weight | 6 | LBW | 10 | 2549.35 | 838.59 | 24602.00 | 1270.49 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | NBW | 10 | 2228.07 | 631.62 | 2343.06 | 914.40 | | | | |
| | 7 | LBW | 12 | 2456.75 | 819.14 | 2587.03 | 1186.94 | 47.00 | -0.86 | 0.418 | 0.22 |
| | | NBW | 10 | 2169.92 | 326.24 | 2227.34 | 474.30 | | | | |
| | 8 | LBW | 10 | 2515.89 | 929.76 | 2292.08 | 1304.89 | 55.00 | -0.33 | 0.771 | 0.08 |
| | | NBW | 12 | 2300.63 | 540.34 | 2163.50 | 838.13 | | | | |
| | 9 | LBW | 10 | 2537.04 | 630.80 | 2486.68 | 686.08 | 46.00 | -0.30 | 0.796 | 0.08 |
| | | NBW | 10 | 2535.48 | 545.71 | 2848.92 | 871.39 | | | | |
| | 10 | LBW | 10 | 2513.19 | 794.71 | 2577.80 | 1160.48 | 43.00 | -0.85 | 0.426 | 0.22 |
| | | NBW | 11 | 2317.81 | 356.89 | 2302.71 | 480.68 | | | | |
| | 11 | LBW | 10 | 2775.10 | 880.73 | 2723.08 | 767.71 | 33.00 | -1.29 | 0.218 | 0.34 |
| | | NBW | 10 | 2410.68 | 397.52 | 2285.31 | 476.74 | | | | |
| | 12 | LBW | 9 | 2340.59 | 780.16 | 2283.80 | 1471.76 | 23.00 | -1.01 | 0.350 | 0.54 |
| | | NBW | 11 | 2708.42 | 670.40 | 2815.01 | 1160.88 | | | | |
| complications | 6 | no | 13 | 2245.38 | 602.46 | 2216.42 | 637.03 | 31.00 | -1.15 | 0.275 | 0.32 |
| | | yes | 7 | 2654.90 | 941.38 | 2450.51 | 1635.96 | | | | |
| | 7 | no | 13 | 2194.31 | 518.15 | 2258.39 | 883.40 | 47.00 | -0.77 | 0.471 | 0.20 |
| | | yes | 9 | 2517.13 | 791.48 | 2524.83 | 991.37 | | | | |
| | 8 | no | 15 | 2310.06 | 588.82 | 2239.92 | 878.91 | 49.00 | -0.25 | 0.837 | 0.07 |
| | | yes | 7 | 2587.95 | 1001.83 | 2159.49 | 1777.05 | | | | |
| | 9 | no | 13 | 2543.99 | 63.56 | 2543.03 | 1057.93 | 45.00 | -0.04 | 0.968 | 0.01 |
| | | yes | 7 | 2521.92 | 368.55 | 2430.33 | 654.47 | | | | |
| | 10 | no | 14 | 2397.00 | 719.89 | 2476.19 | 1087.75 | 46.00 | -0.22 | 0.856 | 0.06 |
| | | yes | 7 | 2438.54 | 262.18 | 2302.71 | 367.15 | | | | |
| | 11 | no | 13 | 2530.70 | 832.11 | 2262.21 | 670.03 | 28.00 | -1.39 | 0.183 | 0.38 |
| | | yes | 7 | 2412.36 | 320.00 | 2677.85 | 279.21 | | | | |
| | 12 | no | 12 | 2526.57 | 685.38 | 2452.26 | 974.42 | 25.00 | -0.53 | 0.646 | 0.17 |
| | | yes | 5 | 2703.60 | 834.46 | 2826.80 | 1588.80 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| care | 6 | bedside | 13 | 2257.53 | 689.31 | 2322.74 | 923.14 | 39.00 | -0.52 | 0.643 | 0.14 |
| | | ward | 7 | 2632.33 | 826.11 | 2275.00 | 998.92 | | | | |
| | 7 | bedside | 13 | 2240.52 | 440.58 | 2258.39 | 701.93 | 51.00 | -0.50 | 0.647 | 0.13 |
| | | ward | 9 | 2450.38 | 882.05 | 2582.38 | 1143.21 | | | | |
| | 8 | bedside | 15 | 2245.14 | 526.29 | 2159.49 | 823.95 | 39.00 | -0.95 | 0.368 | 0.26 |
| | | ward | 7 | 2727.06 | 1021.30 | 2630.78 | 1777.05 | | | | |
| | 9 | bedside | 13 | 2481.58 | 664.62 | 2226.54 | 934.37 | 37.00 | 0.50 | 0.536 | 0.19 |
| | | ward | 7 | 2637.81 | 376.34 | 2543.03 | 411.04 | | | | |
| | 10 | bedside | 14 | 2250.71 | 430.27 | 2292.20 | 713.55 | 26.00 | -1.72 | 0.094 | 0.47 |
| | | ward | 7 | 2731.13 | 782.50 | 2918.10 | 795.88 | | | | |
| | 11 | bedside | 13 | 2399.09 | 474.57 | 2313.64 | 654.76 | 30.00 | -1.23 | 0.241 | 0.34 |
| | | ward | 7 | 2952.80 | 910.09 | 2677.85 | 1223.41 | | | | |
| | 12 | bedside | 14 | 2657.68 | 688.65 | 2699.42 | 1243.30 | 15.00 | -0.76 | 0.509 | 0.29 |
| | | ward | 3 | 2209.52 | 837.11 | 2320.70 | . | | | | |

Appendix 3. Key child daily durations and demographic background variables in monthly inspection

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| birth order | 6 | A | 10 | 2417.83 | 940.64 | 2035.21 | 1705.81 | 43.00 | -0.53 | 0.631 | 0.14 |
| | | B | 10 | 2359.59 | 522.33 | 2316.11. | 243.00 | | | | |
| | 7 | A | 11 | 2470.84 | 774.60 | 2285.83 | 920.43 | 48.00 | -0.821 | 0.44 | 0.21 |
| | | B | 11 | 2181.91 | 482.53 | 2258.39 | 761.96 | | | | |
| | 8 | A | 11 | 2398.59 | 760.58 | 2167.51 | 770.75 | 60.00 | -0.03 | 0.97 | 0.01 |
| | | B | 11 | 2398.36 | 738.89 | 2159.49 | 1128.73 | | | | |
| | 9 | A | 10 | 2434.82 | 696.53 | 2328.43 | 990.31 | 40.00 | -0.76 | 0.45 | 0.20 |
| | | B | 10 | 2637.71 | 433.23 | 2596.40 | 741.56 | | | | |
| | 10 | A | 11 | 2360.15 | 522.07 | 2439.54 | 799.98 | 52.00 | -0.21 | 0.83 | 0.05 |
| | | B | 10 | 2466.61 | 697.03 | 2267.96 | 827.27 | | | | |
| | 11 | A | 10 | 2635.15 | 872.96 | 2629.07 | 697.33 | 46.00 | -0.30 | 0.76 | 0.08 |
| | | B | 10 | 2550.63 | 491.38 | 2287.92 | 748.35 | | | | |
| | 12 | A | 8 | 2609.98 | 596.33 | 2699.42 | 1071.89 | 36.00 | 0.00 | 1.00 | 0.00 |
| | | B | 9 | 2550.70 | 833.57 | 2320.70 | 1437.55 | | | | |
| gender | 6 | girl | 12 | 2645.71 | 2316.11 | 2316.11 | 1340.62 | 29.00 | -1.47 | 0.157 | 0.40 |
| | | boy | 8 | 2003.21 | 443.46 | 2072.54 | 852.69 | | | | |
| | **7** | **girl** | **14** | **2521.15** | **692.99** | **2587.03** | **936.20** | **25.00** | **-2.12** | **0.04** | **0.55** |
| | | **boy** | **8** | **1985.52** | **393.52** | **2044.26** | **630.94** | | | | |
| | 8 | girl | 13 | 2600.06 | 773.03 | 2630.78 | 829.26 | 34.00 | -1.64 | 0.11 | 0.42 |
| | | boy | 9 | 2107.31 | 589.74 | 1958.55 | 417.99 | | | | |
| | 9 | girl | 12 | 2543.80 | 662.48 | 2406.06 | 1081.46 | 47.00 | -0.08 | 0.97 | 0.02 |
| | | boy | 8 | 2524.96 | 452.31 | 2592.83 | 806.17 | | | | |
| | 10 | girl | 12 | 2466.78 | 9.89 | 2524.80 | 842.69 | 48.00 | -0.43 | 0.70 | 0.11 |
| | | boy | 9 | 2336.27 | 436.84 | 2302.71 | 758.31 | | | | |
| | 11 | girl | 12 | 2690.97 | 823.99 | 2612.91 | 617.20 | 36.00 | -0.93 | 0.38 | 0.25 |
| | | boy | 8 | 2445.76 | 432.62 | 2339.92 | 573.34 | | | | |
| | 12 | girl | 11 | 2403.96 | 658.97 | 2320.70 | 1199.90 | 20.00 | -1.31 | 0.22 | 0.39 |
| | | boy | 6 | 2898.75 | 744.31 | 2857.79 | 1463.63 | | | | |
| siblings | 6 | yes | 10 | 2205.27 | 587.69 | 2267.66 | 881.77 | 41.00 | -0.68 | 0.529 | 0.18 |
| | | no | 10 | 2572.16 | 859.86 | 2292.81 | 1295.15 | | | | |
| | 7 | yes | 10 | 2325.92 | 447.77 | 2274.11 | 710.07 | 55.00 | -0.33 | 0.77 | 0.08 |
| | | no | 12 | 2326.75 | 796.76 | 2107.97 | 1040.71 | | | | |
| | 8 | yes | 10 | 2095.63 | 318.26 | 2119.56 | 336.41 | 38.00 | -1.45 | 0.16 | 0.37 |
| | | no | 12 | 2650.86 | 886.84 | 2664.64 | 1578.59 | | | | |
| | 9 | yes | 10 | 2569.89 | 695.51 | 2649.18 | 1060.00 | 45.00 | -0.38 | 0.74 | 0.10 |
| | | no | 10 | 2502.63 | 457.68 | 2406.93 | 635.78 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **10** | **yes** | **10** | **2108.78** | **528.10** | **2098.88** | **1035.40** | **27.00** | **-1.97** | **0.05** | **0.51** |
| | | **no** | **11** | **2685.46** | **539.04** | **2576.94** | **690.73** | | | | |
| | 11 | yes | 10 | 2451.02 | 533.01 | 2446.96 | 689.67 | 49.00 | -0.08 | 0.97 | 0.02 |
| | | no | 10 | 2734.75 | 823.80 | 2457.09 | 779.19 | | | | |
| | 12 | yes | 11 | 2519.61 | 774.78 | 2296.11 | 1183.00 | 29.00 | -0.40 | 0.73 | 0.12 |
| | | no | 6 | 2686.73 | 624.76 | 2699.42 | 959.64 | | | | |
| mom edu | 6 | lower | 10 | 2180.46 | 635.45 | 2126.52 | 925.16 | 37.00 | -0.94 | 0.353 | 0.26 |
| | | higher | 10 | 2596.97 | 812.01 | 2316.11 | 1138.14 | | | | |
| | 7 | lower | 10 | 2376.59 | 464.32 | 2405.33 | 846.18 | 50.00 | -0.659 | 0.54 | 0.17 |
| | | higher | 12 | 2284.70 | 786.32 | 2094.25 | 1049.88 | | | | |
| | 8 | lower | 12 | 2245.10 | 557.40 | 2148.86 | 752.94 | 49.00 | -0.73 | 0.50 | 0.18 |
| | | higher | 10 | 2582.53 | 894.79 | 2538.36 | 1244.43 | | | | |
| | 9 | lower | 10 | 2311.26 | 557.69 | 2220.52 | 783.22 | 29.00 | -1.59 | 0.12 | 0.42 |
| | | higher | 10 | 2761.26 | 521.72 | 2702.55 | 1054.49 | | | | |
| | 10 | lower | 11 | 2246.87 | 441.60 | 2302.71 | 480.68 | 38.00 | -1.20 | 0.25 | 0.31 |
| | | higher | 10 | 2591.22 | 713.86 | 2657.79 | 1071.92 | | | | |
| | 11 | lower | 10 | 2419.20 | 505.42 | 2314.21 | 665.89 | 42.00 | -0.61 | 0.58 | 0.16 |
| | | higher | 10 | 2766.58 | 827.64 | 2612.91 | 773.77 | | | | |
| | 12 | lower | 11 | 2531.27 | 800.61 | 2583.83 | 1346.15 | 28.00 | -0.50 | 0.66 | 0.15 |
| | | higher | 6 | 2665.36 | 562.73 | 2653.09 | 901.89 | | | | |
| dad edu | 6 | lower | 10 | 2515.18 | 832.73 | 2407.20 | 854.27 | 33.00 | -1.29 | 0.218 | 0.34 |
| | | higher | 10 | 2262.24 | 656.16 | 2214.75 | 714.98 | | | | |
| | 7 | lower | 12 | 2553.71 | 682.59 | 2405.33 | 896.96 | 34.00 | -1.714 | 0.09 | 0.43 |
| | | higher | 10 | 2053.57 | 505.31 | 1878.72 | 976.49 | | | | |
| | 8 | lower | 10 | 2379.45 | 889.37 | 2103.14 | 756.88 | 47.00 | -0.86 | 0.42 | 0.22 |
| | | higher | 12 | 2414.33 | 611.94 | 2538.36 | 839.86 | | | | |
| | 9 | lower | 10 | 2401.34 | 420.27 | 2247.81 | 687.57 | 33.00 | -1.29 | 0.22 | 0.34 |
| | | higher | 10 | 2671.19 | 691.83 | 2596.40 | 1068.13 | | | | |
| | 10 | lower | 11 | 2299.12 | 318.55 | 2302.71 | 351.74 | 42.00 | -0.92 | 0.39 | 0.24 |
| | | higher | 10 | 2533.75 | 806.54 | 2743.06 | 1160.48 | | | | |
| | 11 | lower | 10 | 2564.86 | 379.17 | 2564.12 | 531.69 | 40.00 | -0.76 | 0.48 | 0.20 |
| | | higher | 10 | 2620.92 | 928.42 | 2236.22 | 816.50 | | | | |
| | 12 | lower | 9 | 2510.32 | 652.67 | 2296.11 | 1019.47 | 31.00 | -0.81 | 0.67 | 0.14 |
| | | higher | 8 | 2655.40 | 807.63 | 2900.25 | 1453.16 | | | | |

Appendix 4. Female adult daily durations and neonatal health-related background variables in monthly inspection

| | Age | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pretermity | **6** | **fullterm** | **12** | **2464.91** | **791.23** | **2343.71** | **1100.04** | **11** | **-2.855** | **0.003** | **0.77** |
| | | **preterm** | **8** | **1371.8** | **506.88** | **1154.53** | **876.3** | | | | |
| | 7 | fullterm | 14 | 1755.33 | 864.13 | 1702.99 | 1535.76 | 30 | -1.775 | 0.082 | 0.46 |
| | | preterm | 8 | 2603.29 | 1016.48 | 2523.28 | 2047.77 | | | | |
| | 8 | fullterm | 14 | 2287.9 | 863.44 | 2337.53 | 813.25 | 51 | -0.341 | 0.764 | 0.09 |
| | | preterm | 8 | 2124.36 | 790.2 | 2269.55 | 1550.44 | | | | |
| | 9 | full-term | 12 | 2597.69 | 1457.73 | 2485.18 | 2159.62 | 31 | -1.312 | 0.208 | 0.35 |
| | | preterm | 8 | 1718.95 | 887.48 | 1619.21 | 1697.22 | | | | |
| | 10 | fullterm | 13 | 1995.9 | 953.6 | 1881.14 | 1048.62 | 44 | -0.579 | 0.595 | 0.15 |
| | | preterm | 8 | 2151.46 | 1029.7 | 2218.41 | 1783.41 | | | | |
| | 11 | fullterm | 12 | 2443.95 | 893.86 | 2374.11 | 1388.9 | 37 | -0.849 | 0.427 | 0.23 |
| | | preterm | 8 | 2641.7 | 549.42 | 2613.68 | 884.19 | | | | |
| | 12 | fullterm | 10 | 2792.11 | 759.31 | 2957.48 | 995.29 | 21 | -1.366 | 0.193 | 0.40 |
| | | preterm | 7 | 3269.81 | 544.37 | 3209.8 | 611.97 | | | | |
| birth weight | 6 | LBW | 10 | 1856.34 | 700.63 | 1985.03 | 1220.66 | 44 | -0.454 | 0.684 | 0.12 |
| | | NBW | 10 | 2199 | 1022.82 | 1933.38 | 1737.95 | | | | |
| | 7 | LBW | 12 | 1717.71 | 918.08 | 1468.8 | 1601.24 | 31 | -1.912 | 0.059 | 0.48 |
| | | NBW | 10 | 2478.84 | 952.71 | 2304.61 | 1320.71 | | | | |
| | 8 | LBW | 10 | 2333.17 | 949.15 | 2499.11 | 1040.26 | 54 | -0.396 | 0.722 | 0.10 |
| | | NBW | 12 | 2141.16 | 731.75 | 2289.56 | 1053.94 | | | | |
| | 9 | LBW | 10 | 1959.05 | 1618.35 | 1619.21 | 1967.41 | 29 | -1.587 | 0.123 | 0.42 |
| | | NBW | 10 | 2533.33 | 905.17 | 2710.76 | 970.22 | | | | |
| | **10** | **LBW** | **10** | **1429.24** | **707.59** | **1428.66** | **1453.84** | **15** | **-2.817** | **0.004** | **0.73** |
| | | **NBW** | **11** | **2083.86** | **804.27** | **2443.93** | **1216.21** | | | | |
| | 11 | LBW | 10 | 2335.39 | 692.69 | 2333.35 | 1104.7 | 36 | -1.058 | 0.315 | 0.28 |
| | | NBW | 10 | 2710.72 | 821.02 | 2724.84 | 947.87 | | | | |
| | 12 | LBW | 9 | 3295.68 | 823.11 | 3580.1 | 964.79 | 15 | -1.809 | 0.078 | 0.70 |
| | | NBW | 11 | 2821.43 | 603.8 | 2984.29 | 695.24 | | | | |
| complications | 6 | no | 13 | 2147.72 | 815.71 | 2035.01 | 1360.85 | 33 | -0.991 | 0.351 | 0.27 |
| | | yes | 7 | 1804.72 | 991.69 | 1598.88 | 1025.53 | | | | |
| | 7 | no | 13 | 1941.31 | 852.31 | 2048.43 | 1424.48 | 49 | -0.634 | 0.556 | 0.162 |
| | | yes | 9 | 2240.42 | 1194.11 | 2261.52 | 2457.08 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 8 | no | 15 | 2127.76 | 864.76 | 2257.12 | 1000.84 | 33 | -1.375 | 0.185 | 0.37 |
| | | yes | 7 | 2444.17 | 736.2 | 2785.03 | 918.14 | | | | |
| | 9 | no | 13 | 2520.15 | 1324.69 | 2216.42 | 1889.05 | 32 | -1.07 | 0.311 | 0.30 |
| | | yes | 7 | 1737.42 | 1207.56 | 1533.61 | 2429.95 | | | | |
| | 10 | no | 14 | 2112.9 | 823.46 | 2013.68 | 673.72 | 48 | -0.075 | 0.971 | 0.02 |
| | | yes | 7 | 1939.69 | 1258.24 | 2160.85 | 2364.41 | | | | |
| | 11 | no | 13 | 2528.02 | 653.28 | 2533.57 | 1081.95 | 40 | -0.436 | 0.699 | 0.12 |
| | | yes | 7 | 2513.83 | 997.28 | 2514.06 | 990.98 | | | | |
| | 12 | no | 12 | 2958.49 | 796.77 | 3085.04 | 1009.9 | 27 | -0.316 | 0.799 | 0.10 |
| | | yes | 5 | 3061.59 | 470.57 | 3209.8 | 752.67 | | | | |
| care | 6 | bedside | 13 | 2071.88 | 989.38 | 1831.74 | 1540.96 | 44 | -0.119 | 0.938 | 0.03 |
| care | | ward | 7 | 1945.56 | 658.46 | 2066.59 | 101.56 | | | | |
| care | 7 | bedside | 13 | 2322.5 | 993.32 | 2286.6 | 1255.25 | 35 | -1.569 | 0.126 | 0.40 |
| care | | ward | 9 | 1689.82 | 910.45 | 1406.65 | 1221.25 | | | | |
| care | 8 | bedside | 15 | 2094.71 | 952.21 | 2257.13 | 1825.14 | 40 | -0.881 | 0.407 | 0.24 |
| care | | ward | 7 | 2515 | 332.22 | 2594.01 | 650.14 | | | | |
| care | 9 | bedside | 13 | 2516.49 | 1163.57 | 2582.37 | 1359.92 | 24 | -1.704 | 0.097 | 0.47 |
| care | | ward | 7 | 1744.22 | 1507.43 | 1533.61 | 1315.75 | | | | |
| care | **10** | **bedside** | **14** | **2462.53** | **792.18** | **2282.67** | **1071.54** | **13** | **-2.686** | **0.006** | **0.73** |
| care | | **ward** | **7** | **1240.42** | **744.56** | **795.35** | **1465.74** | | | | |
| care | 11 | bedside | 13 | 2582.87 | 779.75 | 2533.57 | 1005.75 | 40 | -0.436 | 0.699 | 0.12 |
| care | | ward | 7 | 2411.97 | 781.08 | 2306.04 | 1614.31 | | | | |
| care | **12** | **bedside** | **14** | **2820.74** | **654.5** | **2985.32** | **749.74** | **0** | **-2.646** | **0.003** | **n.a.** |
| care | | **ward** | **3** | **3773.14** | **219.81** | **3701.2** | **.** | | | | |

Appendix 5. Female adult daily durations and demographic background variables in monthly inspection

| | Age | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| birth order | 6 | A | 10 | 1926.64 | 926.18 | 1694.12 | 1611.54 | 41.00 | -0.68 | 0.530 | 0.18 |
| | | B | 10 | 2128.70 | 849.58 | 2068.63 | 1193.42 | | | | |
| | 7 | A | 11 | 2081.84 | 1073.90 | 2048.43 | 1807.67 | 59.00 | -0.10 | 0.949 | 0.02 |
| | | B | 11 | 2045.52 | 951.86 | 2261.52 | 1518.50 | | | | |
| | 8 | A | 11 | 2184.56 | 899.85 | 2322.00 | 1098.57 | 52.00 | -0.56 | 0.606 | 0.14 |
| | | B | 11 | 2272.31 | 778.42 | 2353.06 | 749.65 | | | | |
| | 9 | A | 10 | 2297.47 | 1365.88 | 2052.59 | 2042.42 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | B | 10 | 2194.92 | 1322.92 | 2143.59 | 1789.96 | | | | |
| | 10 | A | 11 | 2132.30 | 1049.54 | 2136.24 | 1439.28 | 52.00 | -0.21 | 0.863 | 0.05 |
| | | B | 10 | 1970.31 | 900.70 | 2020.99 | 1216.16 | | | | |
| | 11 | A | 10 | 2405.28 | 702.35 | 2447.11 | 1361.80 | 45.00 | -0.38 | 0.739 | 0.10 |
| | | B | 10 | 2640.83 | 841.45 | 2568.89 | 1043.68 | | | | |
| | 12 | A | 8 | 2967.29 | 625.39 | 3186.65 | 894.54 | 33.00 | -0.29 | 0.815 | 0.08 |
| | | B | 9 | 3007.94 | 803.26 | 2986.35 | 1019.47 | | | | |
| gender | 6 | girl | 12 | 2040.50 | 838.12 | 2050.80 | 1155.83 | 44.00 | -0.31 | 0.792 | 0.08 |
| | | boy | 8 | 2008.42 | 977.52 | 1715.31 | 1706.09 | | | | |
| | **7** | **girl** | **14** | **1621.27** | **807.61** | **1468.80** | **1431.76** | **16.00** | **-2.73** | **0.005** | **0.71** |
| | | **boy** | **8** | **2837.90** | **809.40** | **2876.20** | **1498.43** | | | | |
| | 8 | girl | 13 | 2191.99 | 1029.87 | 2496.28 | 1881.01 | 53.00 | -0.37 | 0.744 | 0.09 |
| | | boy | 9 | 2281.08 | 422.67 | 2257.13 | 485.02 | | | | |
| | 9 | girl | 12 | 2165.70 | 1630.30 | 1754.04 | 3032.89 | 37.00 | -0.85 | 0.427 | 0.23 |
| | | boy | 8 | 2366.93 | 672.25 | 2485.18 | 1226.06 | | | | |
| | 10 | girl | 12 | 1800.01 | 637.50 | 2013.68 | 864.22 | 37.00 | -1.03 | 0.247 | 0.31 |
| | | boy | 9 | 2395.36 | 1232.92 | 2443.93 | 2107.09 | | | | |
| | 11 | girl | 12 | 2492.69 | 900.93 | 2437.36 | 1458.38 | 40.00 | -0.62 | 0.571 | 0.17 |
| | | boy | 8 | 2568.60 | 552.06 | 2724.84 | 1082.36 | | | | |
| | 12 | girl | 11 | 2983.50 | 850.59 | 3187.50 | 1288.29 | 26.00 | -0.70 | 0.525 | 0.21 |
| | | boy | 6 | 2998.55 | 366.22 | 2957.47 | 593.15 | | | | |
| siblings | 6 | yes | 10 | 1911.10 | 836.12 | 1816.94 | 1653.44 | 43.00 | -0.53 | 0.631 | 0.14 |
| | | no | 10 | 2144.24 | 934.34 | 1985.03 | 1238.16 | | | | |
| | 7 | yes | 10 | 2194.76 | 1068.51 | 2261.18 | 1389.65 | 51.00 | -0.59 | 0.582 | 0.15 |
| | | no | 12 | 1954.44 | 953.93 | 1570.55 | 1715.04 | | | | |
| | 8 | yes | 10 | 1970.95 | 1156.04 | 1839.37 | 2008.45 | 48.00 | -0.79 | 0.456 | 0.20 |
| | | no | 12 | 2443.01 | 294.76 | 2378.63 | 534.67 | | | | |
| | 9 | yes | 10 | 2535.92 | 1469.85 | 2399.40 | 2245.03 | 38.00 | -0.91 | 0.393 | 0.24 |
| | | no | 10 | 1956.47 | 1128.98 | 1754.04 | 1856.90 | | | | |

| | Group | N | Age | M | SD | Mdn | *IQR* | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | yes | 10 | 2148.75 | 727.94 | 1886.13 | 771.14 | 54.00 | -0.07 | 0.973 | 0.02 |
| | | no | 11 | 1970.08 | 1162.66 | 2289.37 | 1777.68 | | | | |
| | 11 | yes | 10 | 2469.74 | 560.03 | 2578.64 | 643.22 | 49.00 | -0.08 | 0.971 | 0.02 |
| | | no | 10 | 2576.37 | 955.10 | 2260.35 | 3127.77 | | | | |
| | 12 | yes | 11 | 2848.69 | 770.27 | 2986.35 | 1251.83 | 23.00 | -1.01 | 0.350 | 0.30 |
| | | no | 6 | 3245.71 | 521.82 | 3205.40 | 877.10 | | | | |
| mom edu | 6 | lower | 10 | 2100.05 | 1111.33 | 1715.31 | 2112.11 | 49.00 | -0.08 | 0.971 | 0.02 |
| | | higher | 10 | 1955.30 | 596.31 | 2050.80 | 1029.21 | | | | |
| | **7** | **lower** | **10** | **2648.78** | **738.71** | **2493.00** | **1133.79** | **20.00** | **-2.64** | **0.007** | **0.67** |
| | | **higher** | **12** | **1576.10** | **925.05** | **1369.79** | **1435.61** | | | | |
| | 8 | lower | 12 | 2429.02 | 849.27 | 2424.67 | 751.13 | 36.00 | -1.58 | 0.123 | 0.40 |
| | | higher | 10 | 1987.74 | 760.34 | 2159.10 | 1328.16 | | | | |
| | **9** | **lower** | **10** | **2946.00** | **1442.60** | **2939.12** | **2235.75** | **17.00** | **-2.50** | **0.011** | **0.66** |
| | | **higher** | **10** | **1546.39** | **672.48** | **1619.21** | **899.28** | | | | |
| | **10** | **lower** | **11** | **2632.25** | **818.41** | **2443.93** | **936.50** | **11.00** | **-3.10** | **0.001** | **0.80** |
| | | **higher** | **10** | **1420.36** | **671.91** | **1620.90** | **1227.59** | | | | |
| | 11 | lower | 10 | 2698.53 | 692.19 | 2578.64 | 485.25 | 38.00 | -0.91 | 0.393 | 0.24 |
| | | higher | 10 | 2347.58 | 827.41 | 2223.97 | 1677.38 | | | | |
| | 12 | lower | 11 | 3126.32 | 429.73 | 3187.50 | 631.22 | 26.00 | -0.70 | 0.525 | 0.21 |
| | | higher | 6 | 2736.72 | 1046.00 | 2878.92 | 2089.20 | | | | |
| dad edu | 6 | lower | 10 | 2171.64 | 1076.36 | 2084.42 | 2061.98 | 44.00 | -0.45 | 0.684 | 0.12 |
| | | higher | 10 | 1883.70 | 630.62 | 1867.60 | 1046.59 | | | | |
| | 7 | lower | 12 | 2064.05 | 938.38 | 2142.09 | 875.89 | 59.00 | -0.07 | 0.974 | 0.02 |
| | | higher | 10 | 2063.23 | 1101.23 | 2229.61 | 1987.30 | | | | |
| | 8 | lower | 10 | 2203.67 | 794.45 | 2458.01 | 1468.65 | 52.00 | -0.53 | 0.628 | 0.13 |
| | | higher | 12 | 2249.07 | 879.43 | 2305.09 | 635.12 | | | | |
| | 9 | lower | 10 | 1851.55 | 1106.53 | 2302.21 | 2211.59 | 37.00 | -0.98 | 0.353 | 0.26 |
| | | higher | 10 | 2640.84 | 1432.17 | 1846.01 | 2690.03 | | | | |
| | 10 | lower | 11 | 2354.36 | 1130.75 | 2275.96 | 1297.92 | 36.00 | -1.34 | 0.197 | 0.35 |
| | | higher | 10 | 1726.04 | 632.47 | 1851.91 | 1047.95 | | | | |
| | 11 | lower | 10 | 2743.59 | 804.90 | 2724.84 | 758.85 | 34.00 | -1.21 | 0.247 | 0.32 |
| | | higher | 10 | 2302.52 | 690.10 | 2260.35 | 1104.70 | | | | |
| | 12 | lower | 9 | 2974.01 | 390.56 | 2986.35 | 572.49 | 28.00 | -0.77 | 0.481 | 0.22 |
| | | higher | 8 | 3005.46 | 977.02 | 3374.68 | 1662.61 | | | | |

Appendix 6. Male adult daily durations and neonatal health-related background variables in monthly inspection

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pretermity | 6 | full-term | 12 | 852.96 | 547.21 | 742.55 | 785.03 | 41.00 | -0.54 | 0.624 | 0.15 |
| | | preterm | 8 | 896.17 | 313.18 | 897.25 | 574.24 | | | | |
| | 7 | full-term | 14 | 616.66 | 415.98 | 493.15 | 311.04 | 54.00 | -0.14 | 0.920 | 0.04 |
| | | preterm | 8 | 545.24 | 219.94 | 546.26 | 348.31 | | | | |
| | 8 | full-term | 14 | 1281.58 | 1138.69 | 939.89 | 1204.41 | 48.00 | -0.55 | 0.616 | 0.14 |
| | | preterm | 8 | 1378.02 | 775.45 | 1438.93 | 1535.46 | | | | |
| | 9 | full-term | 12 | 824.90 | 314.24 | 812.68 | 335.28 | 24.00 | -1.85 | 0.069 | 0.50 |
| | | preterm | 8 | 1226.83 | 382.32 | 1296.05 | 707.38 | | | | |
| | 10 | full-term | 13 | 1472.86 | 1059.77 | 915.14 | 1924.26 | 37.00 | -1.09 | 0.301 | 0.29 |
| | | preterm | 8 | 854.70 | 360.14 | 802.88 | 679.09 | | | | |
| | 11 | full-term | 12 | 966.12 | 477.25 | 920.68 | 650.80 | 28.00 | -1.54 | 0.135 | 0.42 |
| | | preterm | 8 | 1181.67 | 440.44 | 1272.69 | 464.02 | | | | |
| | 12 | full-term | 10 | 1651.35 | 856.41 | 1384.18 | 1245.18 | 34.00 | -0.10 | 0.962 | 0.03 |
| | | preterm | 7 | 1465.96 | 574.25 | 1438.98 | 298.90 | | | | |
| birth weight | 6 | LBW | 10 | 995.62 | 508.62 | 1040.39 | 675.58 | 34.00 | -1.21 | 0.247 | 0.32 |
| | | NBW | 10 | 744.87 | 386.83 | 651.55 | 389.32 | | | | |
| | 7 | LBW | 12 | 669.56 | 454.27 | 466.90 | 532.52 | 54.00 | -0.40 | 0.722 | 0.10 |
| | | NBW | 10 | 496.03 | 141.40 | 498.76 | 289.88 | | | | |
| | 8 | LBW | 10 | 1462.74 | 132.76 | 1165.47 | 1776.67 | 59.00 | -0.07 | 0.974 | 0.02 |
| | | NBW | 12 | 1194.91 | 703.67 | 1287.23 | 1192.95 | | | | |
| | 9 | LBW | 10 | 877.16 | 325.74 | 832.88 | 702.53 | 35.00 | -1.13 | 0.280 | 0.30 |
| | | NBW | 10 | 1094.18 | 433.89 | 1076.55 | 808.30 | | | | |
| | 10 | LBW | 10 | 885.06 | 728.48 | 757.88 | 612.00 | 29.00 | -1.83 | 0.072 | 0.47 |
| | | NBW | 11 | 1557.65 | 955.61 | 1377.43 | 1820.20 | | | | |
| | 11 | LBW | 10 | 1057.32 | 426.28 | 1087.86 | 686.14 | 47.00 | -0.23 | 0.853 | 0.06 |
| | | NBW | 10 | 1047.36 | 521.80 | 1121.32 | 820.52 | | | | |
| | 12 | LBW | 6 | 1824.14 | 902.33 | 1443.12 | 1387.51 | 25.00 | -0.80 | 0.462 | 0.24 |
| | | NBW | 11 | 1439.12 | 638.32 | 1438.98 | 1175.07 | | | | |
| complications | 6 | no | 13 | 997.26 | 488.40 | 1019.35 | 644.76 | 24.00 | -1.70 | 0.097 | 0.47 |
| | | yes | 7 | 634.36 | 296.55 | 568.47 | 355.49 | | | | |
| | 7 | no | 13 | 654.95 | 432.76 | 518.81 | 461.48 | 49.00 | -0.63 | 0.556 | 0.16 |
| | | yes | 9 | 497.86 | 171.57 | 478.67 | 318.24 | | | | |
| | 8 | no | 15 | 1342.61 | 1119.98 | 1368.20 | 1350.45 | 51.00 | -0.11 | 0.945 | 0.03 |
| | | yes | 7 | 1261.02 | 765.94 | 962.75 | 1472.06 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 9 | no | 13 | 988.10 | 316.63 | 864.24 | 537.24 | 39.00 | -0.52 | 0.643 | 0.14 |
| | | yes | 7 | 981.15 | 529.39 | 709.89 | 1107.61 | | | | |
| | 10 | no | 14 | 1370.53 | 924.02 | 926.92 | 1282.44 | 31.00 | -1.34 | 0.197 | 0.37 |
| | | yes | 7 | 971.05 | 859.60 | 563.26 | 1086.89 | | | | |
| | 11 | no | 13 | 991.40 | 381.96 | 1014.97 | 766.74 | 32.00 | -1.07 | 0.311 | 0.30 |
| | | yes | 7 | 1165.52 | 605.22 | 1239.76 | 908.27 | | | | |
| | 12 | no | 12 | 1536.35 | 761.16 | 1293.67 | 914.29 | 25.00 | -0.53 | 0.646 | 0.17 |
| | | yes | 5 | 1536.35 | 761.16 | 1293.67 | 914.29 | | | | |
| care | 6 | bedside | 13 | 925.42 | 511.93 | 771.29 | 724.08 | 40.00 | -0.44 | 0.699 | 0.12 |
| | | ward | 7 | 767.76 | 349.13 | 775.14 | 603.28 | | | | |
| | 7 | bedside | 13 | 577.07 | 385.18 | 478.70 | 300.33 | 54.00 | -0.30 | 0.794 | 0.08 |
| | | ward | 9 | 610.35 | 321.68 | 507.59 | 501.99 | | | | |
| | 8 | bedside | 15 | 1349.04 | 1038.01 | 1368.20 | 1273.92 | 48.00 | -0.32 | 0.783 | 0.09 |
| | | ward | 7 | 1247.24 | 998.23 | 962.75 | 1445.89 | | | | |
| | 9 | bedside | 13 | 1076.02 | 390.20 | 886.01 | 624.48 | 24.00 | -1.70 | 0.097 | 0.47 |
| | | ward | 7 | 817.88 | 354.62 | 709.89 | 757.17 | | | | |
| | **10** | **bedside** | **14** | **1534.11** | **951.87** | **1250.59** | **1866.70** | **18.00** | **-2.31** | **0.020** | **0.63** |
| | | **ward** | **7** | **643.89** | **365.31** | **507.14** | **648.17** | | | | |
| | 11 | bedside | 13 | 1085.96 | 471.02 | 1145.88 | 851.22 | 43.00 | -0.20 | 0.877 | 0.05 |
| | | ward | 7 | 989.91 | 479.94 | 1014.97 | 867.46 | | | | |
| | 12 | bedside | 14 | 1465.24 | 620.89 | 1380.23 | 1107.35 | 13.00 | -1.01 | 0.362 | 0.38 |
| | | ward | 3 | 2087.24 | 1173.37 | 1564.76 | . | | | | |

Appendix 7. Male adult daily durations and demographic background variables in monthly inspection

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| birth order | 6 | A | 10 | 875.06 | 509.06 | 691.82 | 686.53 | 47.00 | -0.23 | 0.853 | 0.06 |
| | | B | 10 | 865.42 | 429.03 | 916.49 | 703.06 | | | | |
| | 7 | A | 11 | 625.68 | 412.42 | 478.70 | 294.69 | 51.00 | -0.62 | 0.562 | 0.16 |
| | | B | 11 | 555.69 | 297.71 | 507.59 | 370.69 | | | | |
| | 8 | A | 11 | 1,368.87 | 1,132.72 | 1,368.20 | 1,273.92 | 57.00 | -0.23 | 0.847 | 0.06 |
| | | B | 11 | 1,264.43 | 906.76 | 1,167.13 | 1,350.45 | | | | |
| | 9 | A | 10 | 1,008.38 | 401.35 | 875.13 | 703.43 | 43.00 | -0.53 | 0.631 | 0.14 |
| | | B | 10 | 962.96 | 397.88 | 829.34 | 706.28 | | | | |
| | 10 | A | 11 | 1,150.36 | 801.72 | 915.14 | 815.66 | 52.00 | -0.21 | 0.863 | 0.05 |
| | | B | 10 | 1,333.08 | 1,036.83 | 1,037.23 | 1,483.04 | | | | |
| | 11 | A | 10 | 972.17 | 471.87 | 1,060.03 | 931.83 | 43.00 | -0.53 | 0.631 | 0.14 |
| | | B | 10 | 1,132.52 | 465.94 | 1,159.98 | 583.76 | | | | |
| | 12 | A | 8 | 1,411.70 | 538.94 | 1,293.67 | 848.03 | 28.00 | -0.77 | 0.481 | 0.22 |
| | | B | 9 | 1,720.17 | 886.34 | 1,564.76 | 1,397.03 | | | | |
| gender | 6 | girl | 12 | 1,021.73 | 517.71 | 1,096.26 | 805.86 | 25.00 | -1.77 | 0.082 | 0.48 |
| | | boy | 8 | 643.01 | 224.45 | 600.82 | 196.30 | | | | |
| | 7 | girl | 14 | 661.17 | 415.90 | 512.20 | 434.95 | 41.00 | -1.02 | 0.330 | 0.27 |
| | | boy | 8 | 46.35 | 157.77 | 438.22 | 339.28 | | | | |
| | 8 | girl | 13 | 1,414.06 | 1,130.52 | 1,368.20 | 1,002.67 | 55.00 | -0.23 | 0.845 | 0.06 |
| | | boy | 9 | 1,175.94 | 827.21 | 1,167.13 | 1,644.21 | | | | |
| | 9 | girl | 12 | 964.23 | 356.76 | 875.13 | 666.92 | 44.00 | -0.31 | 0.792 | 0.08 |
| | | boy | 8 | 1,017.83 | 458.41 | 840.51 | 932.87 | | | | |
| | 10 | girl | 12 | 1,337.87 | 1,051.94 | 926.92 | 1,780.89 | 48.00 | -0.43 | 0.702 | 0.11 |
| | | boy | 9 | 1,103.36 | 691.23 | 821.42 | 1,009.86 | | | | |
| | 11 | girl | 12 | 1,198.81 | 420.89 | 1,191.98 | 608.98 | 26.00 | -1.70 | 0.098 | 0.46 |
| | | boy | 8 | 832.64 | 462.01 | 782.53 | 726.76 | | | | |
| | 12 | girl | 11 | 1,637.00 | 881.20 | 1,321.47 | 1,277.46 | 32.00 | -0.10 | 0.961 | 0.03 |
| | | boy | 6 | 1,455.86 | 406.02 | 1,495.32 | 648.78 | | | | |
| siblings | **6** | **yes** | **10** | **1,116.65** | **473.51** | **1,178.80** | **761.72** | **20.00** | **-2.27** | **0.023** | **0.60** |
| | | **no** | **10** | **623.84** | **290.04** | **600.82** | **502.14** | | | | |
| | 7 | yes | 10 | 658.33 | 486.88 | 472.52 | 484.00 | 59.00 | -0.07 | 0.947 | 0.02 |
| | | no | 12 | 534.31 | 189.91 | 493.15 | 273.90 | | | | |
| | **8** | **yes** | **10** | **1,858.19** | **1,203.13** | **1,666.38** | **1,430.93** | **28.00** | **-2.11** | **0.036** | **0.53** |
| | | **no** | **12** | **865.36** | **496.19** | **824.24** | **963.60** | | | | |
| | 9 | yes | 10 | 1,086.31 | 351.52 | 875.13 | 563.51 | 30.00 | -1.51 | 0.143 | 0.40 |
| | | no | 10 | 885.03 | 417.68 | 658.18 | 808.37 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | yes | 10 | 1,305.31 | 992.60 | 881.92 | 1,061.84 | 49.00 | -0.42 | 0.705 | 0.11 |
| | | no | 11 | 1,175.61 | 855.70 | 938.71 | 1,359.93 | | | | |
| | 11 | yes | 10 | 1,073.90 | 353.95 | 1,128.76 | 568.54 | 47.00 | -0.23 | 0.853 | 0.06 |
| | | no | 10 | 1,030.79 | 572.48 | 1,080.43 | 953.91 | | | | |
| | 12 | yes | 11 | 1,641.50 | 815.26 | 1,438.98 | 1,255.21 | 29.00 | -0.40 | 0.733 | 0.12 |
| | | no | 6 | 1,453.11 | 622.36 | 1,415.31 | 1,268.00 | | | | |
| mom edu | 6 | lower | 10 | 814.32 | 394.16 | 701.47 | 671.92 | 46.00 | -0.30 | 0.796 | 0.08 |
| | | higher | 10 | 926.16 | 530.05 | 773.22 | 618.82 | | | | |
| | 7 | lower | 10 | 725.10 | 447.78 | 600.72 | 345.55 | 36.00 | -1.58 | 0.123 | 0.40 |
| | | higher | 12 | 478.67 | 208.53 | 393.24 | 335.97 | | | | |
| | 8 | lower | 12 | 1,621.96 | 1,165.08 | 1,387.77 | 1,624.95 | 40.00 | -1.32 | 0.203 | 0.33 |
| | | higher | 10 | 950.28 | 642.95 | 939.89 | 1,361.37 | | | | |
| | 9 | lower | 10 | 1,149.27 | 419.97 | 1,296.05 | 771.03 | 26.00 | -1.81 | 0.075 | 0.48 |
| | | higher | 10 | 822.07 | 291.01 | 787.83 | 382.61 | | | | |
| | 10 | lower | 11 | 1,342.59 | 720.71 | 1,123.76 | 972.96 | 36.00 | -1.34 | 0.197 | 0.35 |
| | | higher | 10 | 1,121.63 | 1,096.85 | 674.11 | 1,211.14 | | | | |
| | 11 | lower | 10 | 1,131.88 | 458.42 | 1,153.32 | 569.74 | 43.00 | -0.53 | 0.631 | 0.14 |
| | | higher | 10 | 972.81 | 479.41 | 1,055.86 | 950.76 | | | | |
| | 12 | lower | 11 | 1,733.61 | 872.86 | 1,551.65 | 1,418.99 | 24.00 | -0.93 | 0.404 | 0.27 |
| | | higher | 6 | 1,284.23 | 252.21 | 1,206.68 | 520.02 | | | | |
| dad edu | 6 | lower | 10 | 729.27 | 339.19 | 651.55 | 373.30 | 33.00 | -1.29 | 0.218 | 0.34 |
| | | higher | 10 | 1,011.21 | 532.96 | 1,040.39 | 740.60 | | | | |
| | 7 | lower | 12 | 468.92 | 161.10 | 452.45 | 309.31 | 37.00 | -1.52 | 0.140 | 0.38 |
| | | higher | 10 | 736.80 | 463.95 | 600.72 | 614.29 | | | | |
| | 8 | lower | 10 | 1,391.09 | 585.80 | 1,387.77 | 975.30 | 43.00 | -1.12 | 0.283 | 0.28 |
| | | higher | 12 | 1,254.62 | 1,276.45 | 809.98 | 1,375.79 | | | | |
| | 9 | lower | 10 | 964.91 | 467.04 | 840.51 | 855.14 | 42.00 | -0.61 | 0.579 | 0.16 |
| | | higher | 10 | 1,006.43 | 318.51 | 875.13 | 593.67 | | | | |
| | 10 | lower | 11 | 1,059.13 | 826.79 | 703.27 | 814.16 | 37.00 | -1.27 | 0.223 | 0.33 |
| | | higher | 10 | 1,433.44 | 983.65 | 1,082.23 | 1,272.78 | | | | |
| | 11 | lower | 10 | 1,133.35 | 477.06 | 1,911.98 | 760.06 | 39.00 | -0.83 | 0.436 | 0.22 |
| | | higher | 10 | 971.34 | 460.30 | 994.58 | 948.68 | | | | |
| | 12 | lower | 9 | 1,611.60 | 593.48 | 1,551.65 | 917.99 | 28.00 | -0.77 | 0.481 | 0.22 |
| | | higher | 8 | 1,533.85 | 916.33 | 1,165.10 | 1,194.91 | | | | |

Appendix 8. Other child daily durations and neonatal health-related background variables in monthly inspection

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| preterimity | **6** | **full-term** | **12** | **1637.12** | **581.77** | **1709.72** | **927.56** | **11.00** | **.2.855** | **0.003** | **0.77** |
| | | **preterm** | **8** | **833.49** | **445.97** | **663.88** | **556.89** | | | | |
| | **7** | **full-term** | **14** | **1310.23** | **453.29** | **1314.03** | **546.26** | **24.00** | **-2.18** | **0.029** | **0.57** |
| | | **preterm** | **8** | **792.81** | **492.87** | **535.15** | **863.13** | | | | |
| | 8 | full-term | 14 | 1717.09 | 727.52 | 1666.97 | 1369.16 | 35.00 | -1.43 | 0.165 | 0.38 |
| | | preterm | 8 | 1263.02 | 517.14 | 1147.83 | 1031.24 | | | | |
| | 9 | full-term | 12 | 1950.25 | 861.95 | 1836.19 | 1586.90 | 33.00 | -1.16 | 0.270 | 0.31 |
| | | preterm | 8 | 1423.82 | 402.26 | 1454.12 | 760.20 | | | | |
| | 10 | full-term | 13 | 1911.15 | 749.68 | 1760.86 | 1341.31 | 42.00 | -0.72 | 0.500 | 0.19 |
| | | preterm | 8 | 1554.70 | 430.20 | 1573.57 | 770.32 | | | | |
| | 11 | full-term | 12 | 1848.47 | 619.89 | 1738.62 | 1297.33 | 46.00 | -0.15 | 0.910 | 0.04 |
| | | preterm | 8 | 1779.31 | 383.01 | 1842.42 | 642.99 | | | | |
| | 12 | full-term | 10 | 1767.63 | 675.69 | 1682.55 | 880.54 | 31.00 | -0.39 | 0.740 | 0.11 |
| | | preterm | 7 | 1669.31 | 473.42 | 1555.22 | 531.62 | | | | |
| birth weight | 6 | LBW | 10 | 1412.59 | 773.13 | 1617.71 | 1347.05 | 49.00 | -0.08 | 0.971 | 0.02 |
| | | NBW | 10 | 1218.74 | 542.39 | 1104.13 | 1033.47 | | | | |
| | 7 | LBW | 12 | 1205.96 | 604.21 | 1261.63 | 1139.42 | 52.00 | -0.53 | 0.628 | 0.13 |
| | | NBW | 10 | 1021.42 | 412.73 | 1028.12 | 778.45 | | | | |
| | 8 | LBW | 10 | 1635.96 | 674.05 | 1863.91 | 1351.24 | 51.00 | -0.59 | 0.582 | 0.15 |
| | | NBW | 12 | 1481.96 | 712.48 | 1304.73 | 1058.73 | | | | |
| | 9 | LBW | 10 | 1790.25 | 910.23 | 1363.68 | 1790.18 | 50.00 | 0.00 | 1.000 | 0.00 |
| | | NBW | 10 | 1689.12 | 589.51 | 1590.46 | 1114.68 | | | | |
| | 10 | LBW | 10 | 1835.92 | 893.04 | 1321.56 | 1697.23 | 49.00 | -0.42 | 0.705 | 0.11 |
| | | NBW | 11 | 1720.21 | 377.01 | 1777.13 | 630.47 | | | | |
| | 11 | LBW | 10 | 1918.01 | 545.63 | 1840.87 | 926.07 | 39.00 | -0.83 | 0.436 | 0.22 |
| | | NBW | 10 | 1723.61 | 517.30 | 1810.76 | 954.78 | | | | |
| | 12 | LBW | 6 | 1675.39 | 804.53 | 1365.59 | 1160.37 | 27.00 | -0.60 | 0.591 | 0.18 |
| | | NBW | 11 | 1755.38 | 472.49 | 1669.34 | 340.74 | | | | |
| complications | 6 | no | 13 | 1345.92 | 584.35 | 1569.00 | 1150.05 | 41.00 | -0.36 | 0.757 | 0.10 |
| | | yes | 7 | 1259.48 | 824.98 | 848.01 | 1321.73 | | | | |
| | 7 | no | 13 | 1142.85 | 534.09 | 1344.23 | 967.85 | 53.00 | -0.37 | 0.744 | 0.09 |
| | | yes | 9 | 1092.07 | 535.94 | 1004.47 | 804.07 | | | | |
| | 8 | no | 15 | 1573.88 | 691.02 | 1541.14 | 1203.06 | 49.00 | -0.25 | 0.837 | 0.07 |
| | | yes | 7 | 1505.01 | 718.16 | 1246.81 | 1449.80 | | | | |
| | 9 | no | 13 | 1953.96 | 829.81 | 1908.45 | 1575.73 | 26.00 | -1.55 | 0.135 | 0.43 |
| | | yes | 7 | 1341.73 | 333.41 | 1192.37 | 715.64 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 10 | no | 14 | 1866.12 | 747.53 | 1673.84 | 1308.03 | 46.00 | -0.22 | 0.856 | 0.06 |
| | | yes | 7 | 1593.85 | 420.04 | 1841.33 | 770.33 | | | | |
| | 11 | no | 13 | 1873.72 | 624.82 | 1925.77 | 1282.97 | 38.00 | -0.59 | 0.588 | 0.16 |
| | | yes | 7 | 1722.55 | 285.18 | 1755.98 | 404.69 | | | | |
| | 12 | no | 12 | 1714.73 | 649.32 | 1612.28 | 803.13 | 29.00 | -0.11 | 0.959 | 0.03 |
| | | yes | 5 | 1756.94 | 460.84 | 1594.46 | 662.83 | | | | |
| care | 6 | bedside | 13 | 1345.74 | 534.57 | 1242.36 | 1045.28 | 35.00 | -0.83 | 0.438 | 0.23 |
| | | ward | 7 | 1259.82 | 889.98 | 689.29 | 1377.37 | | | | |
| | 7 | bedside | 13 | 1175.09 | 484.01 | 1344.23 | 854.66 | 46.00 | -0.84 | 0.431 | 0.21 |
| | | ward | 9 | 1045.51 | 595.21 | 1004.47 | 973.81 | | | | |
| | 8 | bedside | 15 | 1616.97 | 692.44 | 1541.14 | 1122.29 | 45.00 | -0.53 | 0.630 | 0.14 |
| | | ward | 7 | 1412.69 | 694.08 | 1111.03 | 1431.64 | | | | |
| | 9 | bedside | 13 | 1929.93 | 744.96 | 1839.98 | 1231.37 | 24.00 | -1.70 | 0.097 | 0.47 |
| | | ward | 7 | 1386.36 | 662.24 | 1187.36 | 498.24 | | | | |
| | 10 | **bedside** | **14** | **1948.37** | **627.05** | **1845.04** | **585.12** | **20.00** | **-2.16** | **0.031** | **0.59** |
| | | **ward** | **7** | **1429.34** | **619.17** | **1214.71** | **155.74** | | | | |
| | 11 | bedside | 13 | 1887.79 | 574.43 | 1903.31 | 1140.42 | 36.00 | -0.75 | 0.485 | 0.21 |
| | | ward | 7 | 1696.97 | 438.18 | 1637.22 | 831.55 | | | | |
| | 12 | bedside | 14 | 1829.70 | 589.47 | 1682.55 | 635.32 | 7.00 | -1.64 | 0.091 | 0.67 |
| | | ward | 3 | 1248.55 | 266.51 | 1117.28 | . | | | | |

Appendix 9. Other child daily durations and demographic background variables in monthly inspection

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| birth order | 6 | A | 10 | 1292.55 | 645.45 | 1356.12 | 1333.63 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | B | 10 | 1338.78 | 703.57 | 1153.69 | 1113.59 | | | | |
| | 7 | A | 11 | 1082.29 | 519.79 | 1239.43 | 769.39 | 56.00 | -0.30 | 0.797 | 0.07 |
| | | B | 11 | 1161.87 | 547.54 | 1126.70 | 1042.32 | | | | |
| | 8 | A | 11 | 1546.22 | 766.32 | 1184.64 | 1351.08 | 57.00 | -0.23 | 0.847 | 0.06 |
| | | B | 11 | 1557.72 | 626.82 | 1541.14 | 943.90 | | | | |
| | 9 | A | 10 | 1795.71 | 819.48 | 1590.46 | 1366.67 | 47.00 | -0.23 | 0.853 | 0.06 |
| | | B | 10 | 1683.66 | 709.36 | 1454.12 | 1304.87 | | | | |
| | 10 | A | 11 | 1712.64 | 794.58 | 1336.70 | 772.48 | 38.00 | -1.20 | 0.251 | 0.31 |
| | | B | 10 | 1844.36 | 501.61 | 1809.23 | 776.26 | | | | |
| | 11 | A | 10 | 1845.89 | 549.63 | 1829.60 | 936.09 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | B | 10 | 1795.72 | 531.80 | 1810.76 | 980.53 | | | | |
| | 12 | A | 8 | 1703.97 | 691.69 | 1686.82 | 836.25 | 36.00 | 0.00 | 1.000 | 0.00 |
| | | B | 9 | 1747.74 | 516.77 | 1555.22 | 662.83 | | | | |
| gender | 6 | girl | 12 | 1478.58 | 695.92 | 1617.71 | 1190.39 | 31.00 | -1.31 | 0.208 | 0.35 |
| | | boy | 8 | 1071.29 | 546.63 | 901.88 | 953.00 | | | | |
| | 7 | girl | 14 | 1285.04 | 530.43 | 1334.17 | 769.81 | 30.00 | -1.78 | 0.082 | 0.46 |
| | | boy | 8 | 836.90 | 388.15 | 667.67 | 760.24 | | | | |
| | 8 | girl | 13 | 1618.12 | 628.55 | 1681.94 | 1208.07 | 49.00 | -0.63 | 0.556 | 0.16 |
| | | boy | 9 | 1456.42 | 784.40 | 1246.81 | 1314.02 | | | | |
| | 9 | girl | 12 | 1916.39 | 814.02 | 1687.49 | 1586.90 | 31.00 | -1.31 | 0.208 | 0.35 |
| | | boy | 8 | 1474.62 | 586.73 | 1292.36 | 1123.82 | | | | |
| | 10 | girl | 12 | 1966.46 | 779.00 | 1901.40 | 1496.61 | 36.00 | 1.28 | 0.219 | 0.33 |
| | | boy | 9 | 1520.56 | 348.55 | 1586.83 | 587.17 | | | | |
| | 11 | girl | 12 | 1961.95 | 568.06 | 2054.71 | 1129.10 | 28.00 | -1.54 | 0.135 | 0.42 |
| | | boy | 8 | 1609.10 | 401.68 | 1768.75 | 732.46 | | | | |
| | 12 | girl | 11 | 1530.51 | 529.66 | 1709.72 | 868.23 | 25.00 | -0.80 | 0.462 | 0.24 |
| | | boy | 6 | 1100.82 | 727.79 | 823.82 | 722.92 | | | | |
| siblings | 6 | yes | 10 | 1530.51 | 529.66 | 1709.72 | 868.23 | 20.00 | -1.31 | 0.216 | 0.60 |
| | | no | 10 | 1100.82 | 727.79 | 823.82 | 722.92 | | | | |
| | **7** | **yes** | **10** | **1365.85** | **477.56** | **1433.98** | **438.61** | **26.00** | **-2.24** | **0.025** | **0.57** |
| | | **no** | **12** | **918.94** | **485.48** | **784.51** | **681.64** | | | | |
| | **8** | **yes** | **10** | **1977.87** | **547.09** | **2050.40** | **834.58** | **19.00** | **-2.70** | **0.006** | **0.68** |
| | | **no** | **12** | **1197.06** | **586.44** | **958.59** | **729.73** | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **9** | **yes** | **10** | **2326.34** | **610.29** | **2307.84** | **1072.25** | **2.00** | **-3.63** | **p<.0001** | **0.96** |
| | | **no** | **10** | **1153.02** | **210.82** | **1154.20** | **262.03** | | | | |
| | **10** | **yes** | **10** | **2275.66** | **595.75** | **1984.59** | **1065.34** | **4.00** | **-3.59** | **p<.0001** | **0.93** |
| | | **no** | **11** | **1320.54** | **263.86** | **1217.88** | **155.74** | | | | |
| | **11** | **yes** | **10** | **2208.08** | **354.91** | **2165.81** | **746.53** | **6.00** | **-3.33** | **p<.0001** | **0.88** |
| | | **no** | **10** | **1433.54** | **356.45** | **1340.23** | **542.17** | | | | |
| | 12 | yes | 11 | 1879.63 | 653.61 | 1669.34 | 1125.98 | 20.00 | -1.31 | 0.216 | 0.39 |
| | | no | 6 | 1447.59 | 315.26 | 1547.99 | 651.39 | | | | |
| mom edu | 6 | lower | 10 | 1177.65 | 482.25 | 1104.13 | 906.11 | 42.00 | -0.61 | 0.579 | 0.16 |
| | | higher | 10 | 1453.69 | 798.68 | 1661.00 | 1384.63 | | | | |
| | 7 | lower | 10 | 1137.26 | 563.21 | 1028.12 | 987.37 | 55.00 | -0.33 | 0.771 | 0.08 |
| | | higher | 12 | 1109.42 | 511.29 | 1261.63 | 858.78 | | | | |
| | 8 | lower | 12 | 1389.65 | 545.75 | 1304.73 | 1059.42 | 43.00 | -1.12 | 0.283 | 0.28 |
| | | higher | 10 | 1746.75 | 804.87 | 1886.64 | 1590.85 | | | | |
| | 9 | lower | 10 | 1745.25 | 725.94 | 1590.46 | 972.31 | 47.00 | -0.23 | 0.853 | 0.06 |
| | | higher | 10 | 1734.11 | 809.11 | 1363.68 | 1347.62 | | | | |
| | 10 | lower | 11 | 1914.81 | 613.56 | 1848.75 | 665.31 | 34.00 | -1.48 | 0.152 | 0.38 |
| | | higher | 10 | 1621.97 | 703.58 | 1321.56 | 733.67 | | | | |
| | 11 | lower | 10 | 1755.68 | 550.85 | 1872.42 | 1011.31 | 46.00 | -0.30 | 0.796 | 0.08 |
| | | higher | 10 | 1885.94 | 522.91 | 1797.99 | 846.26 | | | | |
| | 12 | lower | 11 | 1582.47 | 431.85 | 1540.75 | 603.23 | 21.00 | -1.21 | 0.256 | 0.36 |
| | | higher | 6 | 1992.38 | 771.54 | 1753.08 | 1348.66 | | | | |
| dad edu | 6 | lower | 10 | 1479.86 | 695.24 | 1497.68 | 1190.70 | 34.00 | -1.21 | 0.247 | 0.32 |
| | | higher | 10 | 1151.47 | 607.89 | 932.32 | 1106.41 | | | | |
| | 7 | lower | 12 | 1187.70 | 476.90 | 1261.63 | 767.93 | 49.00 | -0.73 | 0.497 | 0.18 |
| | | higher | 10 | 1043.33 | 588.71 | 870.91 | 1012.61 | | | | |
| | 8 | lower | 10 | 1804.86 | 789.67 | 2050.40 | 1400.49 | 41.00 | -1.25 | 0.228 | 0.32 |
| | | higher | 12 | 1341.23 | 523.77 | 1236.84 | 799.18 | | | | |
| | 9 | lower | 10 | 1586.35 | 505.49 | 1562.82 | 849.40 | 42.00 | -0.61 | 0.579 | 0.16 |
| | | higher | 10 | 1893.02 | 934.86 | 1481.76 | 1790.18 | | | | |
| | 10 | lower | 11 | 1636.35 | 324.87 | 1760.86 | 624.78 | 53.00 | -0.14 | 0.918 | 0.04 |
| | | higher | 10 | 1928.27 | 892.39 | 1537.76 | 1683.83 | | | | |
| | 11 | lower | 10 | 1744.13 | 327.47 | 1810.76 | 532.97 | 44.00 | -0.45 | 0.684 | 0.12 |
| | | higher | 10 | 1897.49 | 682.64 | 1985.23 | 1509.26 | | | | |
| | 12 | lower | 9 | 1700.02 | 416.94 | 1669.34 | 436.18 | 33.00 | -0.29 | 0.815 | 0.08 |
| | | higher | 8 | 1757.66 | 764.28 | 1547.99 | 1305.48 | | | | |

Appendix 10. Statistical information for neonatal health-related group differences in AWC counts

| | Age | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pretermity | **6** | **full-term** | **12** | **12377.61** | **3666.74** | **13.201.45** | **6490.52** | **20** | **-2.16** | **0.031** | **0.58** |
| | | **preterm** | **8** | **8795.87** | **2440.71** | **8502.89** | **3591.93** | | | | |
| | 7 | full-term | 14 | 8848.78 | 4431.24 | 8184.74 | 8230.03 | 30 | -1.78 | 0.082 | 0.46 |
| | | preterm | 8 | 12490.88 | 3940.02 | 11179.35 | 8014.44 | | | | |
| | 8 | full-term | 14 | 13943.60 | 7518.16 | 13488.60 | 3798.48 | 56 | 0.00 | 1.000 | 0.00 |
| | | preterm | 8 | 14217.36 | 5276.03 | 12842.38 | 10070.58 | | | | |
| | 9 | full-term | 12 | 13282.98 | 6506.35 | 12484.63 | 10950.04 | 38 | -0.77 | 0.473 | 0.21 |
| | | preterm | 8 | 11884.34 | 4852.55 | 10453.54 | 8571.93 | | | | |
| | 10 | full-term | 13 | 133755.60 | 6716.45 | 14021.76 | 11657.78 | 49 | -0.22 | 0.860 | 0.06 |
| | | preterm | 8 | 12656.29 | 6127.76 | 12612.62 | 11462.46 | | | | |
| | 11 | full-term | 12 | 13592.15 | 4809.90 | 1332.39 | 4299.91 | 29 | -1.47 | 0.157 | 0.40 |
| | | preterm | 8 | 15810.60 | 4183.43 | 16017.04 | 6460.08 | | | | |
| | 12 | full-term | 10 | 18086.63 | 6414.66 | 16771.35 | 10123.22 | 25 | -0.98 | 0.364 | 0.29 |
| | | preterm | 7 | 20006.45 | 2959.26 | 20396.56 | 3497.22 | | | | |
| birth weight | 6 | LBW | 10 | 10777.26 | 2904.59 | 9938.07 | 4895.58 | 48 | -0.15 | 0.912 | 0.04 |
| | | NBW | 10 | 11112.57 | 4405.56 | 10395.89 | 7026.15 | | | | |
| | 7 | LBW | 12 | 9151.50 | 4674.73 | 8111.38 | 7402.24 | 41 | -1.25 | 0.21 | 0.32 |
| | | NBW | 10 | 11399.20 | 4262.53 | 10503.29 | 6527.71 | | | | |
| | 8 | LBW | 10 | 15410.07 | 8634.41 | 14425.89 | 11083.02 | 53 | -0.46 | 0.674 | 0.12 |
| | | NBW | 12 | 12904.05 | 4520.66 | 12539.01 | 3712.72 | | | | |
| | 9 | LBW | 10 | 11135.76 | 6662.12 | 9591.90 | 8629.18 | 33 | -1.29 | 0.218 | 0.34 |
| | | NBW | 10 | 14311.28 | 4594.75 | 13445.01 | 8144.28 | | | | |
| | **10** | **LBW** | **10** | **9368.06** | **5225.41** | **9965.28** | **10153.32** | **17** | **-2.68** | **0.006** | **0.69** |
| | | **NBW** | **11** | **16944.78** | **5135.30** | **18562.51** | **10666.35** | | | | |
| | 11 | LBW | 10 | 13893.30 | 4411.34 | 13728.73 | 5051.01 | 42 | -0.61 | 0.579 | 0.16 |
| | | NBW | 10 | 15065.76 | 4928.43 | 14918.76 | 5934.99 | | | | |
| | 12 | LBW | 10 | 21296.59 | 6591.56 | 22186.00 | 9380.51 | 19 | -1.41 | 0.180 | 0.46 |
| | | NBW | 7 | 17557.45 | 4080.78 | 18231.45 | 5553.79 | | | | |
| complications | 6 | no | 13 | 11928.14 | 3061.85 | 12888.99 | 4730.22 | 21 | -1.94 | 0.056 | 0.54 |
| | | yes | 7 | 9118.93 | 4138.56 | 8317.68 | 2520.37 | | | | |
| | 7 | no | 13 | 9962.35 | 4028.08 | 9375.70 | 4935.81 | 56 | -0.17 | 0.896 | 0.04 |
| | | yes | 9 | 10477.71 | 5423.86 | 10445.22 | 10797.90 | | | | |
| | 8 | no | 15 | 13701.17 | 7558.53 | 11707.46 | 6114.59 | 40 | -0.88 | 0.407 | 0.24 |
| | | yes | 7 | 14775.97 | 4558.48 | 14586.93 | 9843.63 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 9 | no | 13 | 13807.68 | 5283.07 | 12639.55 | 9443.12 | 31 | -1.15 | 0.275 | 0.32 |
| | | yes | 7 | 10710.08 | 6599.14 | 8264.05 | 14407.63 | | | | |
| | 10 | no | 14 | 13992.43 | 5292.63 | 13845.01 | 8079.91 | 41 | -0.60 | 0.585 | 0.16 |
| | | yes | 7 | 12025.58 | 8445.42 | 10995.05 | 17885.69 | | | | |
| | 11 | no | 13 | 14221.38 | 3489.76 | 13865.07 | 4038.69 | 42 | -0.28 | 0.817 | 0.08 |
| | | yes | 7 | 14958.95 | 6480.75 | 15972.46 | 9589.56 | | | | |
| | 12 | no | 12 | 18319.85 | 5867.18 | 18254.36 | 8164.26 | 22 | -0.84 | 0.442 | 0.27 |
| | | yes | 5 | 20214.64 | 3417.51 | 20469.51 | 5467.82 | | | | |
| care | 6 | bedside | 13 | 11277.71 | 3951.50 | 12474.11 | 5936.26 | 41 | -0.36 | 0.757 | 0.10 |
| | | ward | 7 | 10326.86 | 3154.00 | 9005.49 | 5139.02 | | | | |
| | 7 | bedside | 13 | 11214.95 | 4925.94 | 10561.37 | 7509.35 | 37 | -1.44 | 0.151 | 0.37 |
| | | ward | 9 | 8668.41 | 3832.42 | 7862.31 | 6326.66 | | | | |
| | 8 | bedside | 15 | 13475.78 | 7243.35 | 11707.46 | 4679.43 | 40 | -0.88 | 0.407 | 0.24 |
| | | ward | 7 | 15258.94 | 5482.25 | 14630.45 | 9498.75 | | | | |
| | 9 | bedside | 13 | 14132.78 | 4901.64 | 13273.37 | 9045.41 | 26 | -1.55 | 0.135 | 0.43 |
| | | ward | 7 | 10106.33 | 6804.41 | 8264.05 | 9011.26 | | | | |
| | **10** | **bedside** | **14** | **16136.49** | **4962.58** | **16762.52** | **10524.06** | **11** | **-2.84** | **0.003** | **0.78** |
| | | **ward** | **7** | **7737.47** | **5176.90** | **4863.43** | **10114.12** | | | | |
| | 11 | bedside | 13 | 14763.64 | 4339.57 | 14465.85 | 4649.01 | 40 | -0.44 | 0.699 | 0.12 |
| | | ward | 7 | 13951.90 | 5348.47 | 12991.62 | 12422.34 | | | | |
| | **12** | **bedside** | **14** | **17576.15** | **443.35** | **18254.36** | **5922.58** | **4** | **-2.14** | **0.032** | **0.81** |
| | | **ward** | **3** | **24948.45** | **4912.35** | **24293.38** | **.** | | | | |

Appendix 11. Statistical information for AWC group differences for demographic variables

| Variable | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| birth order | 6 | A | 10 | 10492.20 | 3657.26 | 9938.07 | 6779.38 | 44.00 | -0.45 | 0.684 | 0.12 |
| | | B | 10 | 11397.63 | 3751.94 | 10581.10 | 5943.26 | | | | |
| | 7 | A | 11 | 10372.62 | 5194.01 | 8407.27 | 7960.80 | 59.00 | -0.10 | 0.949 | 0.02 |
| | | B | 11 | 9973.75 | 4010.12 | 10445.22 | 4933.49 | | | | |
| | 8 | A | 11 | 13984.59 | 7631.96 | 7631.96 | 6114.59 | 56.00 | -0.30 | 0.797 | 0.07 |
| | | B | 11 | 14101.71 | 5894.81 | 13606.63 | 8994.80 | | | | |
| | 9 | A | 10 | 12893.56 | 6042.20 | 11306.08 | 11898.88 | 50.00 | 0.00 | 1 | 0.00 |
| | | B | 10 | 12553.48 | 5875.82 | 12997.96 | 9595.87 | | | | |
| | 10 | A | 11 | 13170.75 | 6050.69 | 13668.26 | 9492.94 | 54.00 | -0.07 | 0.973 | 0.02 |
| | | B | 10 | 13519.49 | 7019.12 | 13371.37 | 13673.14 | | | | |
| | 11 | A | 10 | 13596.09 | 4087.48 | 14165.46 | 6509.99 | 43.00 | -0.53 | 0.631 | 0.14 |
| | | B | 10 | 15362.98 | 5105.99 | 14257.06 | 5038.82 | | | | |
| | 12 | A | 8 | 17837.52 | 4089.46 | 18664.10 | 5302.70 | 299.00 | -0.67 | 0.541 | -7.31 |
| | | B | 9 | 19801.26 | 6171.03 | 20469.51 | 9313.61 | | | | |
| gender | 6 | girl | 12 | 11631.43 | 3422.32 | 11841.72 | 5104.12 | 31.00 | -1.31 | 0.208 | 0.35 |
| | | boy | 8 | 9915.14 | 3931.53 | 8201.90 | 6521.47 | | | | |
| | **7** | **girl** | **14** | **8738.75** | **4308.28** | **8111.38** | **6820.36** | **27.00** | **-1.98** | **0.05** | **0.52** |
| | | **boy** | **8** | **12683.44** | **3981.19** | **12469.07** | **8275.75** | | | | |
| | 8 | girl | 13 | 14561.17 | 7991.72 | 14221.33 | 7554.70 | 49.00 | -0.63 | 0.556 | 0.16 |
| | | boy | 9 | 13294.90 | 4403.05 | 11707.46 | 6666.17 | | | | |
| | 9 | girl | 12 | 12337.59 | 6622.34 | 11777.91 | 11234.77 | 45.00 | -0.23 | 0.851 | 0.06 |
| | | boy | 8 | 13302.42 | 4675.19 | 12484.63 | 8998.48 | | | | |
| | 10 | girl | 12 | 12771.83 | 5984.26 | 12612.62 | 7235.08 | 47.00 | -0.50 | 0.651 | 0.13 |
| | | boy | 9 | 14090.12 | 7135.27 | 16463.24 | 13852.39 | | | | |
| | 11 | girl | 12 | 15146.68 | 5287.84 | 14511.33 | 8110.16 | 43.00 | -0.39 | 0.734 | 0.10 |
| | | boy | 8 | 13478.81 | 3374.09 | 13769.11 | 6079.38 | | | | |
| | 12 | girl | 11 | 19266.51 | 6282.61 | 20396.56 | 9249.65 | 27.00 | -0.60 | 0.591 | 0.18 |
| | | boy | 6 | 18163.30 | 2778.99 | 18641.20 | 5592.04 | | | | |
| siblings | 6 | yes | 10 | 11246.99 | 3224.65 | 12681.55 | 5443.65 | 43.00 | -0.53 | 0.631 | 0.14 |
| | | no | 10 | 10642.84 | 4160.16 | 8846.79 | 6848.14 | | | | |
| | 7 | yes | 10 | 11238.38 | 5347.59 | 10677.45 | 10127.62 | 43.00 | -1.12 | 0.283 | 0.28 |
| | | no | 12 | 9285.52 | 3733.04 | 8161.64 | 5519.97 | | | | |
| | 8 | yes | 10 | 15150.08 | 9539.65 | 12232.56 | 13933.12 | 58.00 | -0.07 | 0.974 | 0.03 |
| | | no | 12 | 13120.71 | 2836.94 | 13592.38 | 3805.25 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 9 | yes | 10 | 14351.94 | 5700.48 | 12484.63 | 11426.25 | 34.00 | -1.21 | 0.247 | 0.32 |
| | | no | 10 | 11095.10 | 5719.17 | 11757.67 | 8268.64 | | | | |
| | 10 | yes | 10 | 14094.53 | 5228.00 | 11276.01 | 9626.70 | 48.00 | -0.42 | 0.705 | 0.13 |
| | | no | 11 | 12647.98 | 7437.81 | 14021.76 | 15781.37 | | | | |
| | 11 | yes | 10 | 14363.73 | 2049.92 | 14511.33 | 2678.11 | 44.00 | -0.45 | 0.684 | 0.12 |
| | | no | 10 | 14595.33 | 6346.73 | 12942.45 | 11591.08 | | | | |
| | 12 | yes | 11 | 18449.87 | 5914.11 | 18277.27 | 7028.95 | 26.00 | -0.70 | 0.525 | 0.21 |
| | | no | 6 | 19660.49 | 4064.35 | 19723.75 | 8857.35 | | | | |
| mom edu | 6 | lower | 10 | 10947.71 | 4580.37 | 8699.68 | 8375.89 | 46.00 | -0.30 | 0.796 | 0.08 |
| | | higher | 10 | 10942.12 | 2632.16 | 11634.28 | 4909.53 | | | | |
| | **7** | **lower** | **10** | **13286.84** | **3548.04** | **12848.07** | **7032.16** | **14.00** | **-3.03** | **0.002** | **0.77** |
| | | **higher** | **12** | **7578.47** | **3572.97** | **7648.74** | **5118.16** | | | | |
| | 8 | lower | 12 | 16211.08 | 7309.54 | 13592.38 | 10108.59 | 40.00 | -1.32 | 0.203 | 0.33 |
| | | higher | 10 | 11441.64 | 4924.22 | 12232.56 | 7132.28 | | | | |
| | **9** | **lower** | **10** | **16188.83** | **5575.49** | **17711.99** | **11163.04** | **18.00** | **-2.42** | **0.015** | **0.64** |
| | | **higher** | **10** | **9258.21** | **3650.52** | **9915.91** | **6390.19** | | | | |
| | 10 | lower | 11 | 16063.92 | 4946.00 | 16463.24 | 10444.77 | 28.00 | -1.90 | 0.061 | 0.49 |
| | | higher | 10 | 10337.00 | 6614.24 | 9496.88 | 12041.42 | | | | |
| | 11 | lower | 10 | 15665.06 | 4593.44 | 15406.71 | 3110.17 | 29.00 | -1.59 | 0.123 | 0.42 |
| | | higher | 10 | 13294.00 | 4504.45 | 12942.45 | 5755.10 | | | | |
| | 12 | lower | 11 | 20250.85 | 4657.22 | 20088.50 | 8664.20 | 21.00 | -1.21 | 0.256 | 0.36 |
| | | higher | 6 | 16358.69 | 5698.07 | 16575.08 | 10951.95 | | | | |
| dad edu | 6 | lower | 10 | 10935.82 | 4139.29 | 9043.58 | 6249.85 | 49.00 | -0.08 | 0.971 | 0.02 |
| | | higher | 10 | 10954.01 | 3282.40 | 11841.72 | 6054.73 | | | | |
| | 7 | lower | 12 | 9620.18 | 4413.42 | 8891.48 | 3334.45 | 47.00 | -0.86 | 0.418 | 0.22 |
| | | higher | 10 | 10836.79 | 4820.43 | 11944.17 | 7285.85 | | | | |
| | 8 | lower | 10 | 14024.61 | 3918.45 | 13913.98 | 5289.88 | 50.00 | -0.66 | 0.539 | 0.17 |
| | | higher | 12 | 14058.60 | 8484.41 | 12381.12 | 9004.46 | | | | |
| | 9 | lower | 10 | 11081.91 | 5451.42 | 11285.84 | 7682.50 | 36.00 | -1.06 | 0.315 | 0.28 |
| | | higher | 10 | 14365.14 | 5949.04 | 12681.05 | 12634.09 | | | | |
| | 10 | lower | 11 | 13812.55 | 7207.76 | 11556.99 | 12370.25 | 51.00 | -0.28 | 0.809 | 0.07 |
| | | higher | 10 | 12813.50 | 5631.39 | 13845.01 | 8623.70 | | | | |
| | 11 | lower | 10 | 15505.08 | 4785.57 | 15219.15 | 3177.52 | 34.00 | -1.21 | 0.247 | 0.32 |
| | | higher | 10 | 13453.99 | 4390.24 | 12570.17 | 12950.28 | | | | |
| | 12 | lower | 9 | 19061.29 | 2996.52 | 19050.95 | 4546.92 | 35.00 | -0.10 | 0.963 | 0.03 |
| | | higher | 8 | 18669.98 | 7226.63 | 17853.90 | 12473.77 | | | | |

Appendix 12. Statistical information for CVC group differences for neonatal health-related background variables

| Variable | Age | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pretermity | 6 | full-term | 8 | 1239.29 | 310.69 | 1187.47 | 440.77 | 24.00 | -1.85 | 0.07 | 0.25 |
| | | preterm | 8 | 949.88 | 290.47 | 882.10 | 558.17 | | | | |
| | 7 | full-term | 14 | 117.84 | 278.23 | 1222.88 | 347.45 | 48.00 | -0.61 | 0.57 | 0.14 |
| | | preterm | 8 | 1106.41 | 376.10 | 1053.80 | 619.83 | | | | |
| | 8 | full-term | 14 | 1225.44 | 331.42 | 1330.01 | 610.86 | 37.00 | -1.30 | 0.21 | 0.34 |
| | | preterm | 8 | 1026.53 | 357.54 | 1054.88 | 637.32 | | | | |
| | 9 | full-term | 12 | 1158.23 | 317.00 | 1130.23 | 486.94 | 38.00 | -0.77 | 0.47 | 0.21 |
| | | preterm | 8 | 1326.27 | 342.97 | 1251.13 | 425.99 | | | | |
| | 10 | full-term | 13 | 1225.67 | 284.80 | 1242.48 | 531.11 | 33.00 | -1.38 | 0.19 | 0.37 |
| | | preterm | 8 | 1395.23 | 182.22 | 1334.17 | 278.70 | | | | |
| | **11** | **full-term** | **12** | **1227.21** | **295.62** | **1252.58** | **534.72** | **22.00** | **-2.01** | **0.05** | **0.54** |
| | | **preterm** | **8** | **1666.07** | **658.61** | **1615.77** | **473.07** | | | | |
| | 12 | full-term | 10 | 1591.73 | 717.68 | 1495.55 | 958.22 | 0.31 | -0.39 | 0.74 | 0.99 |
| | | preterm | 7 | 1411.17 | 405.12 | 1256.95 | 752.07 | | | | |
| birth weight | 6 | LBW | 10 | 1148.04 | 410.64 | 1052.02 | 651.13 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | NBW | 10 | 1099.01 | 241.48 | 1176.50 | 310.32 | | | | |
| | 7 | LBW | 12 | 1086.03 | 353.55 | 989.60 | 602.97 | 42.00 | -1.19 | 0.235 | 0.30 |
| | | NBW | 10 | 1232.26 | 243.05 | 1244.16 | 302.20 | | | | |
| | 8 | LBW | 10 | 1014.86 | 407.71 | 951.94 | 853.20 | 37.00 | -1.52 | 0.14 | 0.38 |
| | | NBW | 12 | 1268.32 | 248.21 | 1330.01 | 335.59 | | | | |
| | 9 | LBW | 10 | 1211.06 | 410.13 | 1145.38 | 427.68 | 46.00 | -0.30 | 0.796 | 0.08 |
| | | NBW | 10 | 1239.83 | 246.39 | 1186.97 | 467.70 | | | | |
| | 10 | LBW | 10 | 1244.66 | 312.19 | 1264.73 | 614.76 | 47.00 | -0.56 | 0.605 | 0.15 |
| | | NBW | 11 | 1331.72 | 207.69 | 1330.25 | 349.38 | | | | |
| | 11 | LBW | 10 | 1450.38 | 667.15 | 1374.39 | 650.64 | 49.00 | -0.08 | 0.971 | 0.02 |
| | | NBW | 10 | 1355.13 | 311.09 | 1377.89 | 458.50 | | | | |
| | **12** | **LBW** | **10** | **1180.23** | **368.64** | **1085.28** | **488.55** | **13.00** | **-2.01** | **0.048** | **0.63** |
| | | **NBW** | **7** | **1701.28** | **634.12** | **1562.92** | **805.38** | | | | |
| complications | 6 | no | 13 | 1093.93 | 314.60 | 1007.36 | 453.94 | 36.00 | -0.75 | 0.485 | 0.21 |
| | | yes | 7 | 1178.48 | 372.88 | 1200.74 | 343.45 | | | | |
| | **7** | **no** | **13** | **1024.71** | **265.55** | **995.80** | **465.56** | **26.00** | **-2.17** | **0.03** | **0.56** |
| | | **yes** | **9** | **1337.08** | **286.83** | **1286.04** | **494.02** | | | | |
| | 8 | no | 15 | 1146.23 | 355.49 | 1198.58 | 569.82 | 50.00 | -0.18 | 0.891 | 0.05 |
| | | yes | 7 | 1167.85 | 354.19 | 1155.68 | 691.27 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 9 | no | 13 | 1230.83 | 385.35 | 1143.31 | 489.12 | 43.00 | -0.20 | 0.877 | 0.05 |
| | | yes | 7 | 1215.45 | 216.56 | 1253.33 | 423.86 | | | | |
| | 10 | no | 14 | 1278.52 | 204.06 | 1354.00 | 531.18 | 48.00 | -0.08 | 0.971 | 0.02 |
| | | yes | 7 | 1313.74 | 153.03 | 1255.19 | 333.71 | | | | |
| | 11 | no | 13 | 1346.27 | 616.84 | 1191.41 | 652.37 | 24.00 | -1.70 | 0.097 | 0.47 |
| | | yes | 7 | 1507.66 | 199.00 | 1480.83 | 380.43 | | | | |
| | 12 | no | 12 | 1542.35 | 653.27 | 1404.41 | 771.61 | 29.00 | -0.11 | 0.959 | 0.03 |
| | | yes | 5 | 1457.45 | 508.60 | 1256.95 | 958.36 | | | | |
| care | 6 | bedside | 13 | 1109.53 | 312.58 | 1168.51 | 361.76 | 43.00 | -0.20 | 0.877 | 0.05 |
| | | ward | 7 | 1149.51 | 381.82 | 1174.19 | 525.99 | | | | |
| | 7 | bedside | 13 | 1183.33 | 271.61 | 1236.93 | 425.43 | 48.00 | -0.70 | 0.483 | 0.18 |
| | | ward | 9 | 1107.96 | 372.11 | 995.80 | 566.58 | | | | |
| | 8 | bedside | 15 | 1186.08 | 289.20 | 1198.58 | 472.29 | 46.00 | -0.46 | 0.680 | 0.12 |
| | | ward | 7 | 1082.45 | 465.84 | 1049.26 | 889.63 | | | | |
| | 9 | bedside | 13 | 1185.88 | 319.33 | 1143.31 | 508.21 | 41.00 | -0.36 | 0.757 | 0.10 |
| | | ward | 7 | 1298.91 | 361.14 | 1173.61 | 397.82 | | | | |
| | 10 | bedside | 14 | 1266.47 | 241.41 | 1288.79 | 491.54 | 39.00 | -0.75 | 0.488 | 0.20 |
| | | ward | 7 | 1337.85 | 307.35 | 1377.74 | 358.35 | | | | |
| | 11 | bedside | 13 | 1288.80 | 348.30 | 1274.94 | 614.73 | 35.00 | -0.83 | 0.438 | 0.23 |
| | | ward | 7 | 1614.38 | 705.23 | 1447.00 | 490.79 | | | | |
| | 12 | bedside | 14 | 1622.15 | 605.22 | 1495.55 | 764.00 | 7.00 | -1.76 | 0.091 | 0.67 |
| | | ward | 3 | 1028.47 | 247.01 | 1011.43 | . | | | | |

Appendix 13. Statistical information for CVC group differences for demographic background variables

| Variable | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| birth order | 6 | A | 10 | 112.93 | 417.27 | 969.24 | 626.95 | 46.00 | -0.30 | 0.796 | 0.08 |
| | | B | 10 | 1117.12 | 232.53 | 1179.34 | 352.84 | | | | |
| | 7 | A | 11 | 1200.48 | 324.84 | 1251.38 | 544.81 | 49.00 | -0.76 | 0.478 | 0.19 |
| | | B | 11 | 1104.52 | 302.89 | 1075.62 | 444.17 | | | | |
| | 8 | A | 11 | 1152.96 | 325.48 | 1198.58 | 569.82 | 59.00 | -0.10 | 0.949 | 0.02 |
| | | B | 11 | 1153.26 | 382.73 | 1155.68 | 742.34 | | | | |
| | 9 | A | 10 | 1256.38 | 417.06 | 124.98 | 477.27 | 43.00 | -0.53 | 0.631 | 0.14 |
| | | B | 10 | 1194.51 | 230.88 | 1115.40 | 694.04 | | | | |
| | 10 | A | 11 | 1287.45 | 284.62 | 1286.98 | 520.27 | 54.00 | -0.07 | 0.973 | 0.02 |
| | | B | 10 | 1293.35 | 244.47 | 1334.17 | 325.09 | | | | |
| | 11 | A | 10 | 1498.03 | 663.45 | 1435.55 | 680.07 | 41.00 | -0.68 | 0.529 | 0.18 |
| | | B | 10 | 1307.48 | 294.23 | 1338.61 | 453.12 | | | | |
| | 12 | A | 8 | 1439.58 | 376.75 | 1332.11 | 611.33 | 35.00 | -0.10 | 0.963 | 0.03 |
| | | B | 9 | 1586.54 | 763.99 | 1428.17 | 1049.44 | | | | |
| gender | 6 | girl | 12 | 1204.76 | 339.17 | 1187.47 | 455.84 | 34.00 | -1.08 | 0.305 | 0.29 |
| | | boy | 8 | 1001.67 | 290.59 | 977.99 | 500.10 | | | | |
| | 7 | girl | 14 | 1109.94 | 327.51 | 1099.92 | 496.61 | 45.00 | -0.75 | 0.482 | 0.20 |
| | | boy | 8 | 1226.98 | 282.61 | 1244.16 | 393.12 | | | | |
| | 8 | girl | 13 | 1129.85 | 336.39 | 1060.50 | 691.55 | 51.00 | -0.50 | 0.647 | 0.13 |
| | | boy | 9 | 1186.71 | 378.99 | 1342.91 | 509.09 | | | | |
| | 9 | girl | 12 | 1156.21 | 370.86 | 1115.40 | 212.57 | 29.00 | -1.47 | 0.157 | 0.40 |
| | | boy | 8 | 1329.30 | 241.32 | 1384.15 | 436.97 | | | | |
| | 10 | girl | 12 | 1219.61 | 286.84 | 1248.83 | 504.35 | 35.00 | -1.35 | 0.193 | 0.35 |
| | | boy | 9 | 1384.47 | 194.73 | 1424.43 | 299.39 | | | | |
| | 11 | girl | 12 | 1363.71 | 630.14 | 1288.36 | 656.92 | 31.00 | -1.31 | 0.208 | 0.35 |
| | | boy | 8 | 1461.32 | 269.55 | 1579.79 | 497.59 | | | | |
| | 12 | girl | 11 | 1330.99 | 444.69 | 1256.95 | 865.39 | 17.00 | -1.61 | 0.122 | 0.48 |
| | | boy | 6 | 1859.11 | 734.06 | 1644.71 | 1020.03 | | | | |
| siblings | 6 | yes | 10 | 1126.91 | 347.74 | 1171.35 | 473.94 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | no | 10 | 1120.13 | 327.61 | 1074.68 | 389.08 | | | | |
| | 7 | yes | 10 | 1174.60 | 310.39 | 1180.57 | 534.52 | 55.00 | -0.33 | 0.771 | 0.08 |
| | | no | 12 | 1134.08 | 322.94 | 1142.22 | 412.81 | | | | |
| | 8 | yes | 10 | 1067.16 | 274.76 | 1064.00 | 549.71 | 42.00 | -1.19 | 0.254 | 0.30 |
| | | no | 12 | 1224.74 | 394.19 | 1330.01 | 693.67 | | | | |
| | 9 | yes | 10 | 1197.30 | 329.45 | 1173.89 | 438.78 | 47.00 | -0.23 | 0.853 | 0.06 |
| | | no | 10 | 1253.59 | 345.07 | 1158.46 | 466.69 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **10** | **yes** | **10** | **1101.60** | **214.70** | **1092.86** | **362.04** | **10.00** | **-3.17** | **0.001** | **0.82** |
| | | **no** | **11** | **1461.77** | **159.32** | **1515.37** | **212.20** | | | | |
| | 11 | yes | 10 | 1207.10 | 381.56 | 1122.87 | 757.26 | 29.00 | -1.59 | 0.123 | 0.42 |
| | | no | 10 | 1598.40 | 562.36 | 1508.16 | 349.49 | | | | |
| | 12 | yes | 11 | 1431.43 | 408.01 | 1380.65 | 692.00 | 29 | -0.4 | 0.733 | 0.12 |
| | | no | 6 | 1674.96 | 878.59 | 1423.25 | 1017.77 | | | | |
| mom edu | 6 | lower | 10 | 1017.77 | 233.74 | 1061.40 | 370.14 | 34.00 | -1.21 | 0.247 | 0.32 |
| | | higher | 10 | 1229.28 | 385.73 | 1220.90 | 629.00 | | | | |
| | 7 | lower | 10 | 1214.32 | 250.92 | 1187.80 | 355.65 | 45.00 | -0.99 | 0.346 | 0.25 |
| | | higher | 12 | 1100.98 | 355.03 | 1096.11 | 583.69 | | | | |
| | 8 | lower | 12 | 1145.57 | 326.57 | 1177.13 | 573.67 | 56.00 | -0.26 | 0.921 | 0.07 |
| | | higher | 10 | 1162.16 | 387.23 | 1218.22 | 643.48 | | | | |
| | 9 | lower | 10 | 1104.85 | 308.00 | 1087.12 | 425.19 | 30.00 | -1.51 | 0.143 | 0.40 |
| | | higher | 10 | 1346.04 | 319.68 | 1241.98 | 447.25 | | | | |
| | 10 | lower | 11 | 1282.19 | 272.69 | 1290.60 | 349.38 | 53.00 | -0.14 | 0.918 | 0.04 |
| | | higher | 10 | 1299.13 | 258.86 | 1310.11 | 525.03 | | | | |
| | 11 | lower | 10 | 1316.50 | 350.77 | 1377.89 | 654.41 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | higher | 10 | 1489.01 | 638.22 | 1374.39 | 454.74 | | | | |
| | 12 | lower | 11 | 1555.83 | 724.28 | 1256.95 | 1076.60 | 31.00 | -0.20 | 0.884 | 0.06 |
| | | higher | 6 | 1446.91 | 304.18 | 1404.41 | 542.12 | | | | |
| dad edu | 6 | lower | 10 | 1179.34 | 340.86 | 1224.62 | 368.78 | 34.00 | -1.21 | 0.247 | 0.32 |
| | | higher | 10 | 1067.71 | 324.29 | 939.87 | 362.18 | | | | |
| | **7** | **lower** | **12** | **1363.19** | **224.25** | **1319.75** | **228.70** | **5.00** | **-3.63** | **p<.0001** | **0.92** |
| | | **higher** | **10** | **899.67** | **182.01** | **859.87** | **235.67** | | | | |
| | 8 | lower | 10 | 1236.47 | 268.71 | 1283.77 | 433.14 | 45.00 | -0.99 | 0.346 | 0.25 |
| | | higher | 12 | 1083.64 | 398.51 | 1058.37 | 815.35 | | | | |
| | 9 | lower | 10 | 1190.59 | 235.21 | 1143.86 | 368.87 | 40.00 | -0.76 | 0.481 | 0.20 |
| | | higher | 10 | 1260.30 | 413.88 | 1158.46 | 426.05 | | | | |
| | 10 | lower | 11 | 1320.42 | 179.98 | 1286.98 | 331.27 | 53.00 | -0.14 | 0.918 | 0.04 |
| | | higher | 10 | 1257.08 | 333.87 | 1354.00 | 614.76 | | | | |
| | 11 | lower | 10 | 1355.21 | 277.31 | 1374.39 | 442.80 | 49.00 | -0.08 | 0.971 | 0.02 |
| | | higher | 10 | 1450.30 | 681.90 | 1424.97 | 687.00 | | | | |
| | 12 | lower | 9 | 1457.39 | 389.41 | 1380.65 | 642.49 | 36.00 | 0.00 | 1 | 0.00 |
| | | higher | 8 | 1584.88 | 799.34 | 1355.97 | 1045.57 | | | | |

Appendix 14. Statistical information for differences of CTC and health-related variables

| | Age | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pretermity | 6 | full-term | 8 | 266.13 | 102.96 | 264.08 | 191.18 | 24.00 | -1.85 | 0.07 | 0.25 |
| | | preterm | 8 | 179.42 | 46.47 | 184.55 | 87.59 | | | | |
| | **7** | **full-term** | **14** | **201.92** | **84.21** | **231.34** | **130.67** | **22.00** | **-2.32** | **0.02** | **0.61** |
| | | **preterm** | **8** | **311.24** | **133.16** | **293.71** | **167.49** | | | | |
| | 8 | full-term | 14 | 277.76 | 91.03 | 293.85 | 74.19 | 53.00 | -0.21 | 0.87 | 0.05 |
| | | preterm | 8 | 294.34 | 152.38 | 267.83 | 246.39 | | | | |
| | 9 | full-term | 12 | 259.38 | 146.01 | 234.78 | 286.84 | 41.00 | -0.54 | 0.62 | 0.15 |
| | | preterm | 8 | 290.62 | 121.75 | 267.85 | 211.40 | | | | |
| | 10 | full-term | 13 | 314.45 | 150.59 | 311.02 | 269.26 | 47.00 | -0.36 | 0.75 | 0.10 |
| | | preterm | 8 | 323.57 | 92.71 | 342.02 | 140.35 | | | | |
| | 11 | full-term | 12 | 302.66 | 105.16 | 320.98 | 155.80 | 29.00 | -1.47 | 0.16 | 0.40 |
| | | preterm | 8 | 392.60 | 118.33 | 388.34 | 168.51 | | | | |
| | 12 | full-term | 10 | 472.72 | 183.96 | 415.32 | 187.64 | 29.00 | -0.59 | 0.60 | 0.17 |
| | | preterm | 7 | 444.10 | 134.81 | 379.29 | 229.74 | | | | |
| birth weight | 6 | LBW | 10 | 21.33 | 52.07 | 184-55 | 75.53 | 36.00 | -1.06 | 0.315 | 0.28 |
| | | NBW | 10 | 261.56 | 117.98 | 250.56 | 228.95 | | | | |
| | **7** | **LBW** | **12** | **195.35** | **78.15** | **196.30** | **142.50** | **29.00** | **-2.04** | **0.041** | **0.52** |
| | | **NBW** | **10** | **297.26** | **130.62** | **259.96** | **134.16** | | | | |
| | 8 | LBW | 10 | 255.30 | 130.79 | 237.86 | 159.82 | 34.00 | -1.71 | 0.093 | 0.43 |
| | | NBW | 12 | 307.53 | 96.68 | 313.34 | 77.79 | | | | |
| | **9** | **LBW** | **10** | **199.34** | **95.41** | **187.02** | **150.96** | **18.00** | **-2.42** | **0.015** | **0.64** |
| | | **NBW** | **10** | **344.42** | **131.57** | **379.78** | **248.19** | | | | |
| | **10** | **LBW** | **10** | **231.48** | **116.21** | **216.88** | **193.80** | **17.00** | **-2.68** | **0.006** | **0.69** |
| | | **NBW** | **11** | **396.51** | **82.77** | **366.61** | **151.45** | | | | |
| | 11 | LBW | 10 | 299.82 | 127.76 | 268.83 | 130.73 | 25.00 | -1.89 | 0.063 | 0.50 |
| | | NBW | 10 | 377.45 | 95.14 | 388.12 | 138.72 | | | | |
| | 12 | LBW | 10 | 367.35 | 12.85 | 369.39 | 23.81 | 15.00 | -1.81 | 0.078 | 0.57 |
| | | NBW | 7 | 511.98 | 183.01 | 453.44 | 314.08 | | | | |
| complications | 6 | no | 13 | 236.65 | 88.58 | 231.61 | 119.53 | 38.00 | -0.59 | 0.588 | 0.16 |
| | | yes | 7 | 221.77 | 110.04 | 202.49 | 78.97 | | | | |
| | 7 | no | 13 | 220.30 | 85.10 | 239.79 | 118.72 | 48.00 | -0.70 | 0.512 | 0.18 |
| | | yes | 9 | 272.54 | 148.31 | 247.45 | 190.39 | | | | |
| | 8 | no | 15 | 289.18 | 124.95 | 305.59 | 194.82 | 48.00 | -0.32 | 0.783 | 0.09 |
| | | yes | 7 | 272.24 | 92.84 | 261.62 | 180.42 | | | | |
| | 9 | no | 13 | 279.66 | 128.93 | 253.16 | 238.29 | 42.00 | -0.28 | 0.817 | 0.08 |
| | | yes | 7 | 257.42 | 153.42 | 274.38 | 305.44 | | | | |
| | 10 | no | 14 | 333.27 | 114.84 | 340.04 | 219.74 | 42.00 | -0.52 | 0.636 | 0.14 |
| | | yes | 7 | 287.24 | 158.83 | 315.68 | 271.11 | | | | |

| | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 11 | no | 13 | 331.73 | 116.63 | 328.58 | 132.50 | 40.00 | -0.44 | 0.699 | 0.12 |
| | | yes | 7 | 351.46 | 124.83 | 397.51 | 246.65 | | | | |
| | 12 | no | 12 | 453.93 | 172.19 | 379.44 | 134.25 | 28.00 | -0.21 | 0.879 | 0.07 |
| | | yes | 5 | 477.75 | 149.03 | 423.22 | 280.46 | | | | |
| care | 6 | bedside | 13 | 251.53 | 107.65 | 237.74 | 171.38 | 32.00 | -1.07 | 0.311 | 0.30 |
| | | ward | 7 | 194.85 | 47.90 | 177.08 | 72.86 | | | | |
| | 7 | bedside | 13 | 273.84 | 130.73 | 258.44 | 97.75 | 34.00 | -1.64 | 0.102 | 0.42 |
| | | ward | 9 | 195.20 | 70.17 | 195.32 | 122.70 | | | | |
| | 8 | bedside | 15 | 281.66 | 105.80 | 305.59 | 183.14 | 46.00 | -0.46 | 0.680 | 0.12 |
| | | ward | 7 | 288.35 | 138.28 | 261.32 | 163.84 | | | | |
| | 9 | bedside | 13 | 301.19 | 141.23 | 291.18 | 266.65 | 29.00 | -1.31 | 0.211 | 0.36 |
| | | ward | 7 | 217.44 | 109.30 | 253.16 | 185.63 | | | | |
| | **10** | **bedside** | **14** | **365.20** | **98.28** | **363.78** | **154.11** | **22.00** | **-2.01** | **0.046** | **0.55** |
| | | **ward** | **7** | **223.38** | **137.40** | **169.49** | **279.72** | | | | |
| | 11 | bedside | 13 | 351.48 | 99.30 | 332.00 | 171.71 | 31.00 | -1.15 | 0.275 | 0.32 |
| | | ward | 7 | 314.79 | 149.46 | 285.34 | 165.20 | | | | |
| | 12 | bedside | 14 | 480.87 | 172.06 | 437.13 | 243.25 | 12.00 | -1.13 | 0.300 | 0.43 |
| | | ward | 3 | 367.88 | 16.83 | 375.81 | . | | | | |

Appendix 15. Statistical information about the differences for CVC and demographic variables

| Variable | Age | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| birth order | 6 | A | 10 | 220.57 | 50.57 | 209.48 | 117.99 | 44.00 | -0.45 | 0.684 | 0.12 |
| | | B | 10 | 242.32 | 109.05 | 220.11 | 139.00 | | | | |
| | 7 | A | 11 | 240.16 | 106.44 | 239.79 | 145.88 | 55.00 | -0.36 | 0.748 | 0.09 |
| | | B | 11 | 243.18 | 128.08 | 247.45 | 99.35 | | | | |
| | 8 | A | 11 | 282.98 | 106.43 | 305.59 | 168.83 | 56.00 | -0.30 | 0.797 | 0.07 |
| | | B | 11 | 284.59 | 125.84 | 286.87 | 178.82 | | | | |
| | 9 | A | 10 | 273.77 | 140.82 | 245.39 | 252.82 | 50.00 | 0.00 | 1 | 0.00 |
| | | B | 10 | 269.99 | 135.16 | 257.24 | 234.10 | | | | |
| | 10 | A | 11 | 313.16 | 125.20 | 317.44 | 175.56 | 53.00 | -0.14 | 0.918 | 0.04 |
| | | B | 10 | 323.16 | 139.61 | 339.16 | 274.24 | | | | |
| | 11 | A | 10 | 347.12 | 135.33 | 328.98 | 222.00 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | B | 10 | 330.15 | 101.22 | 346.05 | 176.34 | | | | |
| | 12 | A | 8 | 434.69 | 115.31 | 401.40 | 93.69 | 34.00 | -0.19 | 0.888 | 0.06 |
| | | B | 9 | 484.26 | 197.96 | 379.29 | 280.46 | | | | |
| gender | 6 | girl | 12 | 237.05 | 79.85 | 229.16 | 92.34 | 37.00 | -0.85 | 0.427 | 0.23 |
| | | boy | 8 | 223.03 | 117.55 | 180.63 | 229.01 | | | | |
| | **7** | **girl** | **14** | **194.70** | **86.09** | **196.30** | **130.96** | **16.00** | **-2.73** | **0.005** | **0.71** |
| | | **boy** | **8** | **323.87** | **117.15** | **277.45** | **130.15** | | | | |
| | 8 | girl | 13 | 269.90 | 130.46 | 269.99 | 164.06 | 40.00 | -1.24 | 0.235 | 0.32 |
| | | boy | 9 | 303.85 | 87.54 | 314.89 | 124.97 | | | | |
| | **9** | **girl** | **12** | **215.74** | **115.07** | **187.98** | **159.07** | **19.00** | **-2.24** | **0.025** | **0.60** |
| | | **boy** | **8** | **356.09** | **121.15** | **379.78** | **245.06** | | | | |
| | 10 | girl | 12 | 287.56 | 133.18 | 313.35 | 204.85 | 37.00 | -1.21 | 0.247 | 0.31 |
| | | boy | 9 | 358.42 | 118.20 | 366.61 | 203.19 | | | | |
| | 11 | girl | 12 | 3115.32 | 134.90 | 290.85 | 181.67 | 33.00 | -1.16 | 0.27 | 0.31 |
| | | boy | 8 | 373.61 | 77.45 | 353.66 | 154.01 | | | | |
| | **12** | **girl** | **11** | **403.77** | **124.77** | **370.86** | **31.03** | **9.00** | **-2.41** | **0.015** | **0.73** |
| | | **boy** | **6** | **565.74** | **178.58** | **480.08** | **288.38** | | | | |
| siblings | 6 | yes | 10 | 240.90 | 81.06 | 250.56 | 96.76 | 36.00 | -1.06 | 0.315 | 0.28 |
| | | no | 10 | 221.98 | 108.98 | 173.75 | 117.68 | | | | |
| | 7 | yes | 10 | 261.99 | 150.43 | 259.96 | 175.98 | 51.00 | -0.59 | 0.582 | 0.15 |
| | | no | 12 | 224.74 | 77.54 | 237.09 | 120.12 | | | | |
| | 8 | yes | 10 | 258.75 | 118.35 | 285.41 | 213.71 | 52.00 | -0.53 | 0.628 | 0.13 |
| | | no | 12 | 304.65 | 110.41 | 296.23 | 114.91 | | | | |
| | 9 | yes | 10 | 271.81 | 135.89 | 223.56 | 252.56 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | no | 10 | 271.95 | 140.14 | 267.85 | 242.65 | | | | |
| | 10 | yes | 10 | 295.01 | 72.18 | 313.35 | 146.44 | 39.00 | -1.13 | 0.282 | 0.29 |
| | | no | 11 | 338.76 | 166.07 | 393.07 | 314.85 | | | | |

| | Age | Group | N | M | SD | Mdn | *IQR* | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 11 | yes | 10 | 315.61 | 98.38 | 320.98 | 177.24 | 40.00 | -0.76 | 0.481 | 0.20 |
| | | no | 10 | 361.67 | 133.66 | 354.06 | 204.97 | | | | |
| | 12 | yes | 11 | 427.99 | 121.86 | 379.58 | 148.84 | 25.00 | -0.80 | 0.462 | 0.24 |
| | | no | 6 | 521.34 | 216.63 | 416.36 | 379.11 | | | | |
| mom edu | 6 | lower | 10 | 229.51 | 109.97 | 197.26 | 137.85 | 44.00 | -0.45 | 0.684 | 0.12 |
| | | higher | 10 | 233.38 | 80.90 | 229.28 | 114.88 | | | | |
| | **7** | **lower** | **10** | **305.15** | **119.72** | **257.47** | **142.78** | **27.00** | **-2.18** | **0.03** | **0.55** |
| | | **higher** | **12** | **188.78** | **82.31** | **196.30** | **159.94** | | | | |
| | 8 | lower | 12 | 301.81 | 82.85 | 303.21 | 79.67 | 45.00 | -0.99 | 0.346 | 0.25 |
| | | higher | 10 | 262.16 | 144.37 | 237.86 | 214.70 | | | | |
| | 9 | lower | 10 | 304.96 | 142.11 | 272.17 | 280.62 | 35.00 | -1.13 | 0.28 | 0.30 |
| | | higher | 10 | 238.79 | 124.41 | 227.64 | 248.52 | | | | |
| | **10** | **lower** | **11** | **373.03** | **115.01** | **366.61** | **151.45** | **25.00** | **-2.11** | **0.036** | **0.55** |
| | | **higher** | **10** | **257.31** | **120.32** | **258.94** | **227.70** | | | | |
| | 11 | lower | 10 | 366.36 | 102.17 | 388.12 | 214.11 | 32.00 | -1.36 | 0.19 | 0.36 |
| | | higher | 10 | 310.92 | 128.74 | 306.84 | 140.25 | | | | |
| | 12 | lower | 11 | 503.32 | 179.74 | 423.22 | 314.08 | 19.00 | -1.41 | 0.18 | 0.42 |
| | | higher | 6 | 383.23 | 87.09 | 368.59 | 139.57 | | | | |
| dad edu | 6 | lower | 10 | 264.74 | 111.74 | 232.34 | 200.12 | 31.00 | -1.44 | 0.165 | 0.38 |
| | | higher | 10 | 198.15 | 60.79 | 178.79 | 112.30 | | | | |
| | 7 | lower | 12 | 267.45 | 126.81 | 252.95 | 81.80 | 45.00 | -0.99 | 0.346 | 0.25 |
| | | higher | 10 | 210.73 | 95.99 | 212.78 | 158.47 | | | | |
| | 8 | lower | 10 | 289.54 | 87.94 | 300.88 | 153.94 | 51.00 | -0.59 | 0.582 | 0.15 |
| | | higher | 12 | 278.99 | 135.32 | 293.03 | 184.19 | | | | |
| | 9 | lower | 10 | 269.17 | 146.06 | 253.79 | 263.58 | 48.00 | -0.15 | 0.912 | 0.04 |
| | | higher | 10 | 274.59 | 129.44 | 257.24 | 231.63 | | | | |
| | 10 | lower | 11 | 338.70 | 144.00 | 362.64 | 168.11 | 44.00 | -0.78 | 0.468 | 0.20 |
| | | higher | 10 | 295.07 | 113.22 | 269.94 | 197.79 | | | | |
| | 11 | lower | 10 | 352.76 | 114.28 | 363.05 | 220.76 | 40.00 | -0.76 | 0.481 | 0.20 |
| | | higher | 10 | 324.51 | 123.36 | 314.85 | 145.83 | | | | |
| | 12 | lower | 9 | 462.55 | 112.96 | 451.05 | 175.59 | 29.00 | -0.67 | 0.541 | 0.19 |
| | | higher | 8 | 459.12 | 212.08 | 377.55 | 269.64 | | | | |

Appendix 16. Statistical information about the group comparisons for vocal milestones

| | CDEV | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pretermity | reduplicated babbling | preterm | 8 | 31.3 | 2.4 | 32.0 | 4.0 | 43 | -0.39 | 0.734 | 0.10 |
| | | full-term | 12 | 31.2 | 4.6 | 30.5 | 5.0 | | | | |
| | variegated babbling | preterm | 6 | 51.5 | 14.3 | 50.0 | 15.0 | 11 | -1.12 | 0.310 | 0.39 |
| | | full-term | 6 | 41.0 | 8.4 | 41.0 | 27.0 | | | | |
| | protowords | preterm | 6 | 50.5 | 4.8 | 48.5 | 10.0 | 19.5 | -1.37 | 0.180 | 0.41 |
| | | full-term | 11 | 44.6 | 7.6 | 48.0 | 12.0 | | | | |
| birth weight | reduplicated babbling | LBW | 10 | 31.4 | 2.5 | 32.0 | 4.0 | 38.5 | -0.88 | 0.390 | 0.23 |
| | | NBW | 10 | 31.0 | 4.9 | 29.5 | 17.0 | | | | |
| | variegated babbling | LBW | 6 | 51.0 | 13.7 | 48.0 | 27.0 | 10 | -1.28 | 0.240 | 0.44 |
| | | NBW | 6 | 41.5 | 9.9 | 39.5 | 16.0 | | | | |
| | **protowords** | **LBW** | **7** | **42.6** | **7.7** | **46.0** | **15.0** | **14.5** | **-2.01** | **0.043** | **0.59** |
| | | **NBW** | **10** | **49.5** | **5.5** | **48.5** | **7.0** | | | | |
| complications | reduplicated babbling | no | 13 | 32.1 | 3.8 | 31.0 | 4.0 | 30 | -1.24 | 0.241 | 0.34 |
| | | yes | 7 | 29.6 | 3.5 | 29.0 | 6.0 | | | | |
| | variegated babbling | no | 9 | 47.4 | 12.5 | 45.0 | 23.0 | 10 | -0.65 | 0.600 | 0.26 |
| | | yes | 3 | 42.7 | 14.0 | 42.0 | . | | | | |
| | protowords | no | 12 | 45.8 | 6.3 | 47.5 | 8.0 | 18.5 | -1.22 | 0.234 | 0.38 |
| | | yes | 5 | 48.6 | 9.6 | 48.0 | 16.0 | | | | |
| care | reduplicated babbling | bedside | 13 | 30.8 | 4.5 | 30.0 | 5.0 | 30 | -1.24 | 0.241 | 0.34 |
| | | ward | 7 | 32.0 | 2.0 | 33.0 | 5.0 | | | | |
| | variegated babbling | bedside | 9 | 43.1 | 8.4 | 43.0 | 13.0 | 8 | -1.02 | 0.373 | 0.41 |
| | | ward | 3 | 55.7 | 19.6 | 67.0 | . | | | | |
| | protowords | bedside | 13 | 47.4 | 7.1 | 48.0 | 10.0 | 18 | -0.91 | 0.412 | 0.31 |
| | | ward | 4 | 44.3 | 7.8 | 46.5 | 14.0 | | | | |
| birth order | **reduplicated babbling** | **A** | **10** | **29.0** | **3.0** | **29.5** | **5.0** | **19** | **-2.37** | **0.019** | **0.62** |
| | | **B** | **10** | **33.0** | **3.7** | **33.0** | **4.0** | | | | |
| | variegated babbling | A | 8 | 45.1 | 12.8 | 44.0 | 22.0 | 13.5 | -0.43 | 0.683 | 0.16 |
| | | B | 4 | 48.5 | 13.1 | 45.0 | 24.0 | | | | |
| | protowords | A | 9 | 44.7 | 8.5 | 46.0 | 13.0 | 23.5 | -1.21 | 0.236 | 0.35 |
| | | B | 8 | 48.8 | 5.0 | 48.5 | 5.0 | | | | |
| gender | reduplicated babbling | girl | 12 | 32.3 | 4.0 | 32.0 | 5.0 | 30 | -1.40 | 0.181 | 0.38 |
| | | boy | 8 | 29.6 | 3.1 | 30.0 | 5.0 | | | | |
| | variegated babbling | girl | 8 | 49.9 | 11.4 | 46.5 | 21.0 | 7 | -1.53 | 0.154 | 0.56 |
| | | boy | 4 | 39.0 | 12.4 | 35.0 | 22.0 | | | | |
| | protowords | girl | 11 | 44.3 | 7.2 | 47.0 | 10.0 | 15.5 | -1.77 | 0.078 | 0.53 |
| | | boy | 6 | 51.0 | 5.0 | 50.5 | 9.0 | | | | |

| Variable | CDEV | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| siblings | reduplicated babbling | yes | 10 | 30.7 | 3.3 | 30.0 | 5.0 | 40 | -0.76 | 0.481 | 0.20 |
| | | no | 10 | 31.7 | 4.4 | 31.0 | 4.0 | | | | |
| | variegated babbling | yes | 6 | 47.7 | 5.6 | 46.5 | 10.0 | 12 | -0.96 | 0.394 | 0.33 |
| | | no | 6 | 44.8 | 17.4 | 36.5 | 35.0 | | | | |
| | protowords | yes | 10 | 46.5 | 7.8 | 48.0 | 11.0 | 35 | 0.00 | 1.000 | 0.00 |
| | | no | 7 | 46.9 | 6.7 | 48.0 | 7.0 | | | | |
| mom edu | reduplicated babbling | lower | 10 | 31.2 | 4.5 | 31.0 | 5.0 | 48.5 | 0.91 | 0.912 | 0.03 |
| | | higher | 10 | 31.2 | 3.3 | 30.5 | 5.0 | | | | |
| | variegated babbling | lower | 7 | 42.1 | 9.5 | 42.0 | 15.0 | 10 | -1.22 | 0.268 | 0.43 |
| | | higher | 5 | 52.0 | 14.8 | 48.0 | 34.0 | | | | |
| | **protowords** | **lower** | **10** | **50.5** | **4.5** | **50.0** | **6.0** | **5.5** | **-2.90** | **0.002** | **0.84** |
| | | **higher** | **7** | **41.1** | **6.9** | **44.0** | **14.0** | | | | |
| dad edu | reduplicated babbling | lower | 10 | 30.8 | 4.7 | 29.0 | 5.0 | 37.5 | -0.95 | 0.353 | 0.25 |
| | | higher | 10 | 31.6 | 2.9 | 31.0 | 3.0 | | | | |
| | variegated babbling | lower | 4 | 42.8 | 11.4 | 42.5 | 21.0 | 12 | -0.68 | 0.570 | 0.25 |
| | | higher | 8 | 48.0 | 13.3 | 46.5 | 27.0 | | | | |
| | protowords | lower | 9 | 47.9 | 7.0 | 48.0 | 7.0 | 27 | -0.87 | 0.423 | 0.25 |
| | | higher | 8 | 45.3 | 7.6 | 46.5 | 13.0 | | | | |

Appendix 17. Statistical information about the group comparisons for vocabulary

| | MB-CDI | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| pretermity | 12 receptive | term | 12 | 59.2 | 52.2 | 52.5 | 104.0 | 45 | -0.23 | 0.851 | 0.06 |
| | | preterm | 8 | 51.3 | 23.5 | 43.0 | 44.0 | | | | |
| | 12 expressive | term | 12 | 2.8 | 2.3 | 3.0 | 5.0 | 45 | -0.24 | 0.851 | 0.06 |
| | | preterm | 8 | 3.0 | 2.9 | 1.5 | 8.0 | | | | |
| | **18 expressive** | **term** | **12** | **36.8** | **41.3** | **18.0** | **60.0** | **17** | **-2.54** | **0.016** | **0.65** |
| | | **preterm** | **8** | **38.1** | **24.4** | **36.5** | **45.0** | | | | |
| | 24 expressive | term | 12 | 210.0 | 109.1 | 229.5 | 175.0 | 32 | -1.23 | 0.230 | 0.33 |
| | | preterm | 8 | 265.9 | 77.7 | 268.5 | 136.0 | | | | |
| birth weight | 12 receptive | LBW | 7 | 55.4 | 40.3 | 51.5 | 70.0 | 48 | -0.11 | 0.912 | 0.04 |
| | | NBW | 13 | 56.6 | 46.6 | 39.5 | 69.0 | | | | |
| | **12 expressive** | **LBW** | **7** | **4.1** | **2.6** | **4.0** | **4.0** | **19.5** | **-2.35** | **0.019** | **0.61** |
| | | **NBW** | **13** | **1.6** | **1.8** | **1.0** | **3.0** | | | | |
| | 18 expressive | LBW | 7 | 33.4 | 26.4 | 26.0 | 46.0 | 37 | -1.04 | 0.353 | 0.26 |
| | | NBW | 13 | 41.2 | 42.7 | 22.0 | 55.0 | | | | |
| | 24 expressive | LBW | 7 | 231.3 | 98.4 | 244.0 | 160.0 | 47 | -0.23 | 0.853 | 0.06 |
| | | NBW | 17 | 233.4 | 105.9 | 249.0 | 147.0 | | | | |
| complications | **12 receptive** | **no** | **13** | **70.2** | **41.7** | **76.0** | **79.0** | **19.5** | **-2.06** | **0.037** | **0.57** |
| | | **yes** | **7** | **29.7** | **31.3** | **27.0** | **39.0** | | | | |
| | 12 expressive | no | 13 | 3.2 | 2.7 | 3.0 | 4.0 | 39 | -0.53 | 0.643 | 0.14 |
| | | yes | 7 | 2.3 | 2.1 | 1.0 | 4.0 | | | | |
| | 18 expressive | no | 13 | 43.6 | 39.2 | 31.0 | 61.0 | 35 | -0.88 | 0.438 | 0.23 |
| | | yes | 7 | 25.6 | 22.6 | 21.0 | 42.0 | | | | |
| | 24 expressive | no | 13 | 245.1 | 75.9 | 256.0 | 84.0 | 39 | -0.52 | 0.643 | 0.14 |
| | | yes | 7 | 208.7 | 137.2 | 207.0 | 264.0 | | | | |
| care | 12 receptive | bedisde | 13 | 58.0 | 44.7 | 43.0 | 70.0 | 43 | -0.20 | 0.877 | 0.05 |
| | | ward | 7 | 52.3 | 40.9 | 60.0 | 86.0 | | | | |
| | 12 expressive | bedisde | 13 | 2.2 | 2.2 | 1.0 | 4.0 | 26.5 | -1.53 | 0.135 | 0.42 |
| | | ward | 7 | 4.0 | 2.8 | 4.0 | 4.0 | | | | |
| | 18 expressive | bedisde | 13 | 44.6 | 38.3 | 40.0 | 54.0 | 41 | -0.38 | 0.757 | 0.10 |
| | | ward | 7 | 23.7 | 23.7 | 14.0 | 23.0 | | | | |
| | 24 expressive | bedisde | 13 | 251.1 | 106.0 | 256.0 | 131.0 | 30 | -1.23 | 0.241 | 0.34 |
| | | ward | 7 | 197.6 | 82.0 | 207.0 | 162.0 | | | | |
| birth order | 12 receptive | A | 10 | 53.6 | 43.8 | 39.5 | 63.0 | 48 | -0.15 | 0.912 | 0.04 |
| | | B | 10 | 58.4 | 43.2 | 59.5 | 130.0 | | | | |
| | 12 expressive | A | 10 | 3.1 | 3.0 | 2.5 | 4.0 | 49.5 | -0.04 | 0.971 | 0.01 |
| | | B | 10 | 2.6 | 2.0 | 2.0 | 4.0 | | | | |
| | 18 expressive | A | 10 | 46.8 | 41.9 | 44.0 | 64.0 | 34 | -1.28 | 0.247 | 0.32 |
| | | B | 10 | 27.8 | 24.4 | 20.0 | 27.0 | | | | |
| | 24 expressive | A | 10 | 228.3 | 120.3 | 244.0 | 179.0 | 49 | -0.08 | 0.971 | 0.02 |
| | | B | 10 | 236.4 | 79.9 | 249.0 | 108.0 | | | | |

| Variable | MB-CDI | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| gender | 12 receptive | boy | 8 | 62.9 | 39.4 | 68.0 | 57.0 | 40 | -0.62 | 0.571 | 0.17 |
| | | girl | 12 | 51.4 | 45.4 | 36.5 | 82.0 | | | | |
| | 12 expressive | boy | 8 | 2.6 | 3.3 | 1.0 | 5.0 | 37 | -0.86 | 0.427 | 0.23 |
| | | girl | 12 | 3.0 | 1.9 | 3.0 | 3.0 | | | | |
| | 18 expressive | boy | 8 | 35.3 | 28.1 | 23.0 | 51.0 | 39 | -0.74 | 0.521 | 0.19 |
| | | girl | 12 | 38.7 | 39.7 | 25.0 | 52.0 | | | | |
| | 24 expressive | boy | 8 | 234.8 | 110.1 | 236.5 | 154.0 | 46 | -0.15 | 0.910 | 0.04 |
| | | girl | 12 | 230.6 | 96.9 | 256.0 | 167.0 | | | | |
| siblings | 12 receptive | no | 10 | 54.9 | 52.3 | 45.0 | 94.0 | 45 | -0.38 | 0.739 | 0.10 |
| | | yes | 10 | 57.1 | 32.6 | 43.0 | 54.0 | | | | |
| | 12 expressive | no | 10 | 3.7 | 2.6 | 4.0 | 4.0 | 29.5 | -1.58 | 0.123 | 0.41 |
| | | yes | 10 | 2.0 | 2.2 | 1.0 | 3.0 | | | | |
| | 18 expressive | no | 10 | 35.8 | 45.9 | 13.0 | 70.0 | 45 | -0.40 | 0.739 | 0.10 |
| | | yes | 10 | 38.8 | 21.0 | 35.5 | 36.0 | | | | |
| | **24 expressive** | **no** | **10** | **176.2** | **89.7** | **172.5** | **162.0** | **17** | **-2.50** | **0.011** | **0.66** |
| | | **yes** | **10** | **288.5** | **76.5** | **291.5** | **145.0** | | | | |
| mom edu | 12 receptive | lower | 10 | 64.8 | 50.2 | 43.0 | 97.0 | 38 | -0.91 | 0.393 | 0.24 |
| | | higher | 10 | 47.2 | 33.2 | 45.0 | 58.0 | | | | |
| | 12 expressive | lower | 10 | 2.0 | 1.7 | 1.0 | 3.0 | 34.5 | -1.19 | 0.247 | 0.31 |
| | | higher | 10 | 3.7 | 2.9 | 3.5 | 5.0 | | | | |
| | 18 expressive | lower | 10 | 48.2 | 40.0 | 44.0 | 48.0 | 38 | -0.96 | 0.393 | 0.24 |
| | | higher | 10 | 26.4 | 26.3 | 16.5 | 23.0 | | | | |
| | 24 expressive | lower | 10 | 262.9 | 118.9 | 283.0 | 169.0 | 26 | -1.81 | 0.075 | 0.48 |
| | | higher | 10 | 201.8 | 68.6 | 212.0 | 137.0 | | | | |
| dad edu | **12 receptive** | **lower** | **10** | **32.3** | **27.9** | **31.5** | **45.0** | **18** | **-2.42** | **0.015** | **0.64** |
| | | **higher** | **10** | **79.7** | **42.1** | **85.5** | **90.0** | | | | |
| | **12 expressive** | **lower** | **10** | **1.4** | **1.6** | **1.0** | **2.0** | **15.5** | **-2.66** | **0.007** | **0.69** |
| | | **higher** | **10** | **4.3** | **2.5** | **4.0** | **3.0** | | | | |
| | 18 expressive | lower | 10 | 24.5 | 22.0 | 16.5 | 44.0 | 34 | -1.28 | 0.247 | 0.32 |
| | | higher | 10 | 50.1 | 41.3 | 35.5 | 60.0 | | | | |
| | 24 expressive | lower | 10 | 199.2 | 115.9 | 198.0 | 206.0 | 32 | -1.36 | 0.190 | 0.36 |
| | | higher | 10 | 265.5 | 71.0 | 268.5 | 58.0 | | | | |

Appendix 18. Group comparisons of the MB-CDI 12 month language scores

| | Variable | Group | N | M | SD | Mdn | IQR | $U$ | $Z$ | $p$ | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| First signs of understanding | pretermity | preterm | 8 | 3.00 | 0.00 | 3.00 | 1.0 | 44 | -0.445 | 0.792 | 0.08 |
| | | full-term | 12 | 3.00 | 0.00 | 3.00 | 0.0 | | | | |
| | birth weight | LBW | 7 | 2.86 | 0.38 | 3.00 | 1.0 | 41.5 | -0.457 | 0.757 | 0.09 |
| | | NBW | 13 | 2.77 | 0.44 | 3.00 | 1.0 | | | | |
| | **complications** | **no** | **13** | **3.00** | **0.00** | **3.00** | **.** | **19.5** | **-2.97** | **0.037** | **0.57** |
| | | **yes** | **7** | **2.43** | **0.53** | **2.00** | **1.0** | | | | |
| | care | bedside | 13 | 2.85 | 0.38 | 3.00 | 1.0 | 39.5 | -0.685 | 0.643 | 0.13 |
| | | ward | 7 | 2.71 | 0.49 | 3.00 | 1.0 | | | | |
| | birth order | A | 10 | 2.80 | 0.42 | 3.00 | 1.0 | 50 | 0 | 1 | 0 |
| | | B | 10 | 2.80 | 0.42 | 3.00 | 1.0 | | | | |
| | gender | boy | 8 | 2.75 | 0.46 | 3.00 | 1.0 | 44 | -0.445 | 0.792 | 0.08 |
| | | girl | 12 | 2.83 | 0.39 | 3.00 | 1.0 | | | | |
| | siblings | no | 10 | 2.80 | 0.42 | 3.00 | 0.0 | 50 | 0 | 1 | 0 |
| | | yes | 10 | 2.80 | 0.42 | 3.00 | 1.0 | | | | |
| | mom edu | lower | 10 | 2.80 | 0.42 | 3.00 | 0 | 50 | 0 | 1 | 0 |
| | | higher | 10 | 2.80 | 0.42 | 3.00 | 6.75 | | | | |
| | dad edu | lower | 10 | 2.60 | 0.52 | 3.00 | 1.0 | 30 | -2.179 | 0.143 | 0.4 |
| | | higher | 10 | 3.00 | 0.00 | 3.00 | . | | | | |
| Sum of actions and gestures | pretermity | preterm | 7 | 29.86 | 5.24 | 31.00 | 9.0 | 24 | -0.798 | 0.47 | 0.24 |
| | | full-term | 9 | 27.44 | 4.77 | 28.00 | 5.5 | | | | |
| | birth weight | LBW | 5 | 25.00 | 4.47 | 25.00 | 8.0 | 11 | -1.879 | 0.069 | 0.60 |
| | | NBW | 11 | 30.09 | 4.48 | 29.00 | 7.0 | | | | |
| | complications | no | 10 | 28.40 | 3.66 | 28.00 | 5.25 | 29 | -0.109 | 0.958 | 0.03 |
| | | yes | 6 | 28.67 | 7.06 | 29.00 | 13.0 | | | | |
| | care | bedside | 11 | 30.09 | 4.48 | 29.00 | 7.0 | 11 | -1.879 | 0.069 | 0.60 |
| | | ward | 5 | 25.00 | 4.47 | 25.00 | 8.0 | | | | |
| | birth order | A | 7 | 30.57 | 4.79 | 29.00 | 9.0 | 19 | -1.33 | 0.21 | 0.40 |
| | | B | 9 | 26.89 | 4.73 | 27.00 | 8.0 | | | | |
| | gender | boy | 8 | 27.63 | 4.96 | 26.50 | 6.75 | 21 | -1.161 | 0.279 | 0.56 |
| | | girl | 12 | 29.38 | 5.15 | 30.00 | 6.0 | | | | |
| | **siblings** | **no** | **10** | **25.50** | **3.21** | **26.50** | **4.25** | **8** | **-2.534** | **0.01** | **0.84** |
| | | **yes** | **10** | **31.50** | **4.69** | **31.00** | **6.75** | | | | |
| | **mom edu** | **lower** | **10** | **31.10** | **3.90** | **31.00** | **6.75** | **4** | **-2.835** | **0.003** | **0.87** |
| | | **higher** | **6** | **24.17** | **3.25** | **24.00** | **5.25** | | | | |
| | dad edu | lower | 9 | 28.78 | 5.65 | 28.00 | 7.5 | 27 | -0.479 | 0.68 | 0.14 |
| | | higher | 7 | 28.14 | 4.34 | 27.00 | 6.0 | | | | |

Appendix 19. Group comparisons for MB-CDI language scores at 18 months

| | Variable | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Use of words | pretermity | preterm | 8 | 4.00 | 2.00 | 4.00 | 4.0 | 35.00 | -1.04 | 0.343 | 0.27 |
| | | fullterm | 12 | 3.00 | 3.00 | 2.00 | 2.0 | | | | |
| | birth weight | LBW | 7 | 3.29 | 2.29 | 3.00 | 2.0 | 44.00 | -0.12 | 0.938 | 0.03 |
| | | NBW | 13 | 3.62 | 2.53 | 2.00 | 3.0 | | | | |
| | complications | no | 13 | 3.92 | 2.63 | 3.00 | 4.0 | 32.00 | -1.11 | 0.311 | 0.30 |
| | | yes | 7 | 2.71 | 1.80 | 2.00 | 3.0 | | | | |
| | care | bedside | 13 | 3.77 | 2.49 | 3.00 | 3.0 | 35.00 | -0.82 | 0.438 | 0.23 |
| | | ward | 7 | 3.00 | 2.31 | 2.00 | 1.0 | | | | |
| | birth order | A | 10 | 3.80 | 2.44 | 3.50 | 3.0 | 40.00 | -0.78 | 0.481 | 0.20 |
| | | B | 10 | 3.20 | 2.44 | 2.00 | 2.0 | | | | |
| | gender | boy | 8 | 3.88 | 3.00 | 2.50 | 5.0 | 45.50 | -0.20 | 0.851 | 0.05 |
| | | girl | 12 | 3.25 | 2.01 | 2.50 | 2.0 | | | | |
| | siblings | no | 10 | 3.80 | 3.19 | 2.00 | 6.0 | 43.00 | -0.55 | 0.631 | 0.14 |
| | | yes | 10 | 3.20 | 1.32 | 3.00 | 2.0 | | | | |
| | mom edu | lower | 10 | 4.30 | 2.63 | 4.00 | 5.0 | 28.00 | -1.73 | 0.105 | 0.44 |
| | | higher | 10 | 2.70 | 1.95 | 2.00 | 1.0 | | | | |
| | dad edu | lower | 10 | 2.60 | 1.58 | 2.00 | 5.0 | 27.50 | -1.76 | 0.089 | 0.45 |
| | | higher | 10 | 4.40 | 2.80 | 3.00 | 6.0 | | | | |
| Sum of morphology | **pretermity** | **preterm** | **8** | **1.50** | **1.10** | **1.25** | **2.13** | **11.00** | **-2.99** | **0.003** | **0.77** |
| | | **fullterm** | **12** | **0.21** | **0.26** | **0.00** | **.50** | | | | |
| | birth weight | LBW | 7 | 1.00 | 1.04 | 1.00 | 1.50 | 31.00 | -1.21 | 0.275 | 0.32 |
| | | NBW | 13 | 0.58 | 0.91 | 0.50 | .50 | | | | |
| | complications | no | 13 | 0.69 | 0.83 | 0.50 | 1.00 | 39.00 | -0.54 | 0.643 | 0.14 |
| | | yes | 7 | 0.79 | 1.22 | 0.00 | 2.00 | | | | |
| | care | bedside | 13 | 0.81 | 1.11 | 0.50 | 1.25 | 45.00 | -0.04 | 1 | 0.01 |
| | | ward | 7 | 0.57 | 0.61 | 0.50 | 1.00 | | | | |
| | birth order | A | 10 | 0.90 | 0.99 | 0.50 | 1.63 | 37.00 | -1.03 | 0.353 | 0.26 |
| | | B | 10 | 0.55 | 0.93 | 0.25 | 0.63 | | | | |
| | gender | boy | 8 | 0.56 | 0.78 | 0.25 | 1.25 | 40.00 | -0.65 | 0.571 | 0.17 |
| | | girl | 12 | 0.83 | 1.07 | 0.50 | 1.00 | | | | |
| | siblings | no | 10 | 0.45 | 0.55 | 0.25 | 1.00 | 38.50 | -0.91 | 0.393 | 0.23 |
| | | yes | 10 | 1.00 | 1.20 | 0.50 | 2.25 | | | | |
| | mom edu | lower | 10 | 1.05 | 1.17 | 0.50 | 1.88 | 31.50 | -1.47 | 0.165 | 0.37 |
| | | higher | 10 | 0.40 | 0.57 | 0.00 | 1.00 | | | | |
| | dad edu | lower | 10 | 0.85 | 1.29 | 0.00 | 2.25 | 40.50 | -0.75 | 0.481 | 0.19 |
| | | higher | 10 | 0.60 | 0.46 | 0.50 | 0.63 | | | | |

| | Variable | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MSL | pretermity | preterm | 8 | 8.92 | 0.49 | 9.00 | 0.83 | 42.50 | -0.46 | 0.678 | 0.11 |
| | | fullterm | 12 | 5.33 | 3.74 | 5.00 | 0.61 | | | | |
| | birth weight | LBW | 7 | 1.24 | 0.37 | 1.00 | 0.33 | 38.50 | -0.60 | 0.588 | 0.15 |
| | | NBW | 13 | 1.46 | 0.63 | 1.30 | 0.85 | | | | |
| | complications | no | 13 | 1.36 | 0.41 | 1.33 | 0.52 | 37.00 | -0.72 | 0.536 | 0.19 |
| | | yes | 7 | 1.43 | 0.79 | 1.00 | 1.00 | | | | |
| | care | bedside | 13 | 1.38 | 0.62 | 1.00 | 0.52 | 40.50 | -0.43 | 0.699 | 0.11 |
| | | ward | 7 | 1.38 | 0.45 | 1.33 | 1.00 | | | | |
| | birth order | A | 10 | 1.47 | 0.50 | 1.33 | 1.00 | 36.00 | -1.14 | 0.315 | 0.28 |
| | | B | 10 | 1.30 | 0.62 | 1.00 | 0.33 | | | | |
| | gender | boy | 8 | 1.25 | 0.39 | 1.00 | 0.60 | 36.00 | -0.99 | 0.384 | 0.25 |
| | | girl | 12 | 1.47 | 0.64 | 1.33 | 0.83 | | | | |
| | siblings | no | 10 | 1.36 | 0.53 | 1.00 | 1.00 | 15.50 | -0.37 | 0.739 | 0.69 |
| | | yes | 10 | 1.40 | 0.61 | 1.32 | 0.42 | | | | |
| | mom edu | lower | 10 | 1.40 | 0.69 | 1.00 | 0.57 | 44.00 | -0.49 | 0.684 | 0.12 |
| | | higher | 10 | 1.37 | 0.40 | 1.32 | 1.00 | | | | |
| | dad edu | lower | 10 | 1.40 | 0.66 | 1.00 | 0.77 | 42.50 | -0.61 | 0.579 | 0.15 |
| | | higher | 10 | 1.36 | 0.45 | 1.33 | 0.50 | | | | |

Appendix 20. The group comparisons of the MB-CDI language score at 24 months

| | Variable | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Use of words | pretermity | preterm | 6 | 7.00 | 2.00 | 7.00 | 5.00 | 33.00 | -0.29 | 0.82 | 0.08 |
| | | full-term | 12 | 8.00 | 2.00 | 8.00 | 3.00 | | | | |
| | birth weight | LBW | 5 | 7.80 | 2.59 | 9.00 | 5.00 | 25.00 | -0.75 | 0.503 | 0.23 |
| | | NBW | 13 | 7.31 | 1.65 | 8.00 | 2.00 | | | | |
| | **complications** | **no** | **12** | **8.33** | **1.50** | **9.00** | **2.00** | **6.50** | **-2.81** | **0.003** | **0.82** |
| | | **yes** | **6** | **5.67** | **1.21** | **5.50** | **2.00** | | | | |
| | care | bedside | 13 | 7.23 | 1.83 | 8.00 | 4.00 | 24.50 | -0.80 | 0.443 | 0.25 |
| | | ward | 5 | 8.00 | 2.12 | 8.00 | 4.00 | | | | |
| | birth order | A | 9 | 7.44 | 2.07 | 8.00 | 4.00 | 39.50 | -0.09 | 0.931 | 0.02 |
| | | B | 9 | 7.44 | 1.81 | 8.00 | 4.00 | | | | |
| | gender | boy | 8 | 7.33 | 2.07 | 8.00 | 4.00 | 35.00 | -0.10 | 0.964 | 0.27 |
| | | girl | 12 | 7.50 | 1.88 | 8.00 | 4.00 | | | | |
| | siblings | no | 8 | 7.50 | 2.33 | 8.00 | 5.00 | 36.00 | -0.36 | 0.762 | 0.10 |
| | | yes | 10 | 7.40 | 1.58 | 8.00 | 3.00 | | | | |
| | mom edu | lower | 10 | 6.70 | 1.77 | 6.50 | 3.00 | 18.50 | -1.94 | 0.055 | 0.54 |
| | | higher | 8 | 8.38 | 1.69 | 9.00 | 3.00 | | | | |
| | **dad edu** | **lower** | **10** | **6.30** | **1.70** | **6.00** | **3.00** | **9.00** | **-2.80** | **0.004** | **0.78** |
| | | **higher** | **8** | **8.88** | **0.83** | **9.00** | **2.00** | | | | |
| Sum of morphology | **pretermity** | **preterm** | **6** | **8.92** | **0.49** | **9.00** | **0.38** | **12.00** | **-2.26** | **0.024** | **0.67** |
| | | **full-term** | **12** | **5.33** | **3.74** | **5.00** | **4.63** | | | | |
| | birth weight | LBW | 5 | 8.80 | 2.86 | 9.00 | 4.50 | 17.00 | -1.54 | 0.143 | 0.48 |
| | | NBW | 13 | 5.65 | 3.39 | 5.00 | 4.50 | | | | |
| | complications | no | 12 | 6.58 | 3.57 | 6.25 | 4.63 | 33.00 | -0.28 | 0.82 | 0.08 |
| | | yes | 6 | 6.42 | 3.64 | 7.50 | 5.38 | | | | |
| | care | bedside | 13 | 6.31 | 4.02 | 5.00 | 5.00 | 29.50 | -0.30 | 0.775 | 0.09 |
| | | ward | 5 | 7.10 | 1.60 | 7.50 | 3.00 | | | | |
| | birth order | A | 9 | 7.11 | 4.04 | 8.00 | 5.25 | 32.00 | -0.76 | 0.489 | 0.21 |
| | | B | 9 | 5.94 | 2.94 | 5.00 | 4.38 | | | | |
| | gender | boy | 6 | 5.08 | 3.56 | 5.00 | 6.50 | 24.00 | -1.13 | 0.291 | 0.33 |
| | | girl | 12 | 7.25 | 3.36 | 7.75 | 4.50 | | | | |
| | **siblings** | **no** | **8** | **4.38** | **3.17** | **4.00** | **7.13** | **14.50** | **-2.28** | **0.021** | **0.64** |
| | | **yes** | **10** | **8.25** | **2.77** | **9.00** | **4.88** | | | | |
| | mom edu | lower | 10 | 6.20 | 3.99 | 8.25 | 6.38 | 38.50 | -0.13 | 0.897 | 0.04 |
| | | higher | 8 | 6.94 | 2.93 | 5.50 | 3.75 | | | | |
| | dad edu | lower | 10 | 5.90 | 3.31 | 5.50 | 4.88 | 34.50 | -0.49 | 0.633 | 0.14 |
| | | higher | 8 | 7.31 | 3.76 | 7.75 | 4.75 | | | | |

| | Variable | Group | N | M | SD | Mdn | IQR | *U* | *Z* | *p* | $r_{rb}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MSL | pretermity | preterm | 6 | 5.31 | 1.81 | 5.25 | 3.00 | 20.00 | -1.09 | 0.313 | 0.33 |
| | | full-term | 10 | 4.10 | 2.18 | 3.84 | 2.58 | | | | |
| | birth weight | LBW | 5 | 5.10 | 2.05 | 5.00 | 3.59 | 25.50 | -0.23 | 0.827 | 0.07 |
| | | NBW | 11 | 4.30 | 2.13 | 4.00 | 1.66 | | | | |
| | complications | no | 10 | 4.68 | 1.89 | 5.17 | 1.79 | 25.50 | -0.49 | 0.635 | 0.15 |
| | | yes | 6 | 4.33 | 2.52 | 3.67 | 3.83 | | | | |
| | care | bedside | 11 | 4.74 | 2.44 | 5.33 | 2.33 | 18.00 | -1.09 | 0.32 | 0.35 |
| | | ward | 5 | 4.13 | 0.96 | 3.67 | 1.84 | | | | |
| | birth order | A | 8 | 4.60 | 2.19 | 4.50 | 2.46 | 29.50 | -0.27 | 0.798 | 0.08 |
| | | B | 8 | 4.50 | 2.10 | 4.50 | 1.92 | | | | |
| | gender | boy | 6 | 3.93 | 1.93 | 3.67 | 3.33 | 23.50 | -0.46 | 0.661 | 0.35 |
| | | girl | 12 | 4.83 | 2.16 | 5.00 | 2.17 | | | | |
| | **siblings** | **no** | **6** | **2.89** | **1.59** | **3.33** | **3.00** | **4.00** | **-2.84** | **0.003** | **0.87** |
| | | **yes** | **10** | **5.55** | **1.68** | **5.33** | **2.66** | | | | |
| | mom edu | lower | 8 | 4.52 | 2.52 | 5.33 | 4.21 | 26.50 | -0.58 | 0.574 | 0.17 |
| | | higher | 8 | 4.58 | 1.69 | 3.84 | 1.83 | | | | |
| | dad edu | lower | 10 | 4.15 | 2.24 | 3.67 | 2.88 | 23.00 | -0.77 | 0.492 | 0.23 |
| | | higher | 6 | 5.22 | 1.72 | 5.17 | 2.25 | | | | |

| Vastauspvm. | Nro. |
|---|---|
| | |

TAMPEREEN YLIOPISTO

ESITIETOLOMAKE

Tutkimus: Kaksoslasten jokeltelun ja varhaisen sanaston kehitys

Tutkijan yhteystiedot: Hanna Elo
Tampereen yliopisto, Yhteiskunta- ja kulttuuritieteiden yksikkö
p. 040 1901 333, 040 5200 205
hanna.elo@uta.fi

Tutkimukseen osallistuvien perheiden tiedot käsitellään luottamuksellisesti ja tutkijaa sitoo vaitiolovelvollisuus. Tutkimustulosten raporteissa ei julkaista mitään sellaista tietoa, joka rikkoisi tutkittavien yksityisyydensuojaa eikä tutkittavia voi raporteissa esitettävien tietojen perusteella tunnistaa. Tutkittavien tunnistetiedot ja tutkimukseen liittyvät asiakirjat (suostumusasiakirjat, esitiedot ja muut lomakkeilla kerättävät tiedot) suojataan ja säilytetään analysoitavasta aineistosta erillään mahdollisia jatkoyhteydenottoja varten. Tutkimukseen osallistuvilla perheillä on halutessaan oikeus nähdä arkistoitu materiaali ja kieltää se käyttö tässä tutkimuksessa. Tutkimuksen suostumusasiakirjoihin tutkittavat perheet määrittelevät aineiston jatkokäyttöön liittyvät ehdot.

**Tutkimukseen osallistuvien lasten henkilötiedot**

A-Lapsen nimi: ______________________ B-Lapsen nimi: ______________________
Sukupuoli: ☐M ☐N _Sukupuoli: ☐M ☐N

Lasten syntymäaika: ______________________

**Yhteystiedot**
Vastaavan vanhemman nimi: ______________________
Osoite: ______________________
Puhelinnumero: ______________________
Sähköposti: ______________________

**Lasten kulttuurinen ja kielellinen tausta**

Lasten kielellinen/kulttuurinen tausta:
☐ suomalainen, äidinkieli suomi ☐ suomalainen, äidinkieli ruotsi
☐ muu ______________________

Puhutaanko kotonanne muita kieliä kuin suomea?
Jos, mitä kieliä käytetään? ______________________

**Perhetiedot**

Äidin koulutustaso:
☐ Peruskoulu
☐Ammatillinen opistotutkinto
☐Ylioppilas (IB, EB, Reifeprüfung)
☐ Alempi korkeakoulututkinto
☐ Ylempi korkeakoulututkinto
☐ Muu,
mikä?________________________________
____________________________________

Isän koulutustaso:
☐ Peruskoulu
☐Ammatillinen opistotutkinto
☐Ylioppilas (IB, EB, Reifeprüfung)
☐ Alempi korkeakoulututkinto
☐ Ylempi korkeakoulututkinto
☐ Muu,
mikä?________________________________
____________________________________

Perheen muut lapset

| Ikä | sukupuoli | Mahdolliset puheen/kielen/kuulon pulmat |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |

Lähisuvussa esiintyvät kommunikointiin vaikuttavat tekijät: Esiintyykö lasten lähisukulaisilla jotakin seuraavista (rastita)

Äidin suvussa
☐ puheen-/kielen tai muun kommunikoinnin vaikeuksia
☐ lukemisen, kirjoittamisen tai oppimisen vaikeuksia
☐ kuulovammoja
☐ kehityksellisiä neurologisia poikkeavuuksia (esim. autismin kirjon häiriöitä tai kehitysvammaisuutta),
mitä_________________________________
_____________________________________

Isän suvussa
☐ puheen-/kielen tai muun kommunikoinnin vaikeuksia
☐ lukemisen, kirjoittamisen tai oppimisen vaikeuksia
☐ kuulovammoja
☐ kehityksellisiä neurologisia poikkeavuuksia (esim. autismin kirjon häiriöitä tai kehitysvammaisuutta),
mitä_________________________________
_____________________________________

**Raskaus- ja synnytysaika**

Oliko äidillä sairauksia raskauden aikana? ☐kyllä ☐ei
Tarkempi kuvaus sairauksista:
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________

Käyttikö äiti päihteitä tai tupakoiko äiti raskausaikana? □kyllä □ei □en halua vastata
Tarkempi kuvaus päihteidenkäytöstä ja tupakoinnista: ______________________________

______________________________

______________________________

______________________________

Millä raskausviikolla synnytys alkoi?_______________

Lasten laskettu aika: ______________________________

Lapsen apgar-pisteet A: ______________ B:________________

Syntymäpaino ja –pituus A: ______________ B:________________

Synnytystapa
□ molemmat syntyivät alateitse
□ A alateitse, B sektiolla
□ molemmat sektiolla
- □ suunniteltu sektio
- □ kiireellinen sektio
- □ hätäsektio

Oliko raskaudessa tai synnytyksessä mitään tavallisuudesta poikkeavaa? □kyllä □ei

Tarkempi kuvaus raskauden ja/tai synnytyksen poikkeavuuksista: ______________________________

______________________________

______________________________

______________________________

______________________________

______________________________

______________________________

Lasten hoidontarve vastasyntyneisyyskaudella

A-lapsi
□ vierihoito
□ lastenosastotasoinen hoito
□ tehohoito

B-lapsi
□ vierihoito
□ lastenosastotasoinen hoito
□ tehohoito

Oliko A-lapsella vastasyntyneisyyskaudella nenämahaletkua tai syömisen/nielemisen pulmia? ______________________________

______________________________

______________________________

______________________________

______________________________

______________________________

______________________________

Oliko B-lapsella vastasyntyneisyyskaudella nenämahaletkua tai syömisen/nielemisen pulmia? ______________________________

______________________________

______________________________

______________________________

______________________________

______________________________

______________________________

Kotiutuiko A-lapsi sairaalasta yhtä aikaa äidin kanssa? □kyllä □ei

Jos A-lapsi oli hoidossa lastenosastolla tai teho-osastolla, miksi? Kuinka pitkään lapsi oli hoidossa?

Kotiutuiko B-lapsi sairaalasta yhtä aikaa äidin kanssa? □kyllä □ei

Jos B-lapsi oli hoidossa lastenosastolla tai teho-osastolla, miksi? Kuinka pitkään lapsi oli hoidossa?

**Lasten syöminen ja nieleminen**

Miten lapsenne syövät?

A-lapsi
- □ nenämahaletkulla
- □ osittain nenämahaletkulla, osittain rinnalla
- □ osittain nenämahaletkulla, osittain pullolla
- □ rintaruokinnalla
- □ osittain rinnalla/osittain pullolla
- □ pullolla
- □ muu (esim. PEG-letku)

B-lapsi
- □ nenämahaletkulla
- □ osittain nenämahaletkulla, osittain rinnalla
- □ osittain nenämahaletkulla, osittain pullolla
- □ rintaruokinnalla
- □ osittain rinnalla/osittain pullolla
- □ pullolla
- □ muu (esim. PEG-letku)

Lisätietoa A-lapsen syömisestä ja nielemisestä:

Lisätietoa B-lapsen syömisestä ja nielemisestä:

Käyttäväkö/käyttivätkö lapsesi tuttia?

A-lapsi
□kyllä □ei
Milloin tutti jäi pois? _________

B-lapsi
□kyllä □ei
Milloin tutti jäi pois? _________

Syövätkö lapsesi kiinteitä ruokia?
□kyllä □ei
Missä iässä kiinteät ruoat on aloitettu?________________________________________

**Lasten nykyinen terveydentila ja sairaudet**

Onko lapsillasi ollut mitään seuraavista sairauksista?

A-lapsi
□aivoverenvuoto
□korvatulehduksia
□päähän kohdistunut vamma
□kohtauksia (epilepsia esim.)
□korkeita kuumeita
□hengitystieinfektioita
□univaikeuksia
□aivokalvontulehdus
□ sydämen toiminnan vajavuutta
□allergioita, mitä?
____________________________________
____________________________________
□hengitysvaikeuksia,
mitä?_______________________________
____________________________________
□näköön liittyviä vaikeuksia,
mitä?_______________________________
____________________________________

B-lapsi
□aivoverenvuoto
□korvatulehduksia
□päähän kohdistunut vamma
□kohtauksia (epilepsia esim.)
□korkeita kuumeita
□hengitystieinfektioita
□univaikeuksia
□aivokalvontulehdus
□ sydämen toiminnan vajavuutta
□allergioita, mitä?
____________________________________
____________________________________
□hengitysvaikeuksia,
mitä?_______________________________
____________________________________
□näköön liittyviä vaikeuksia,
mitä?_______________________________
____________________________________

Tarkentavia tietoja edellä mainituista A-lapsen sairauksista, muista sairauksista, vammoista ja/tai leikkauksista
____________________________________
____________________________________
____________________________________
____________________________________

Tarkentavia tietoja edellä mainituista B-lapsen sairauksista, muista sairauksista, vammoista ja/tai leikkauksista
____________________________________
____________________________________
____________________________________
____________________________________

Onko A-lapsella säännöllistä lääkitystä
□kyllä □ei
mihin sairauteen?
____________________________________
____________________________________
____________________________________

Onko B-lapsella säännöllistä lääkitystä
□kyllä □ei
mihin sairauteen?
____________________________________
____________________________________
____________________________________

**Tämänhetkinen ääntelyn kehityksen, vuorovaikutuksen ja kuulon tilanne**

Mitä seuraavista lapsesi tekevät

A-lapsi

☐ itkee
☐ katsoo silmiin
☐ hymyilee
☐ nauraa
☐ ääntelee
☐ jokeltelee yksittäisiä tavuja (esim. (ka)
_ jokeltelee sarjallisesti (esim. baba, annannannaa)
☐ seuraa katseella henkilöitä, leluja
☐ reagoi ääniin
☐ tavoittelee esineitä
☐ muuta, mitä?

______________________________
______________________________
______________________________

B-lapsi

☐ itkee
☐ katsoo silmiin
☐ hymyilee
☐ nauraa
☐ ääntelee
☐ jokeltelee yksittäisiä tavuja (esim. (ka)
_ jokeltelee sarjallisesti (esim. baba, annannannaa)
☐ seuraa katseella henkilöitä, leluja
☐ reagoi ääniin
☐ tavoittelee esineitä
☐ muuta, mitä?

______________________________
______________________________
______________________________

**Muita olennaisia tietoja lapsista**

A-lapsi

______________________________
______________________________
______________________________
______________________________
______________________________
______________________________
______________________________
______________________________
______________________________

B-lapsi

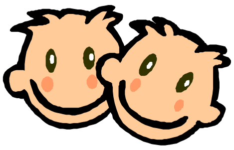
______________________________
______________________________
______________________________
______________________________
______________________________
______________________________
______________________________
______________________________
______________________________

**Tutkimustiedote: Kaksoslasten jokeltelun ja varhaisen sanaston kehitys**

Hei!

Olen kaksoslasten äiti, puheterapeutti ja teen tutkimusta kaksoslasten jokeltelun ja sanaston kehityksestä. Etsin nyt perheitä osallistumaan tutkimukseeni, jotta saisimme tärkeää tietoa kaksoslasten puheen ja kielen kehitykseen vaikuttavista tekijöistä. Tavoittelen nyt niitä **perheitä, joissa lasten laskettu aika on 08/2012 – 12/2012 välisenä aikana.**

**Tutkimuksen tausta:**
Kaksoslasten kielellistä kehitystä verrataan yksöslapsiin ja tässä vertailussa kaksosten kielen kehitys on useammin hitaampaa. Nykyisten tutkimusten perusteella näyttäisi myös, että kaksoslapsilla on riski kielen kehityksen poikkeavuuteen. Tällä hetkellä ei kuitenkaan tiedetä, milloin kaksoslasten hitaampi tai poikkeava kehitys alkaa ja mitä taustatekijöitä eri kehityspoluilla on.



**Tämä tutkimus selvittää**
● Minkälaisia sanojen rakennuspalikoita lapset jokelteluissaan harjoittelevat: millaisia äänteitä ja tavurakenteita lapset tuottavat
● Mitä sanoja lapset käyttävät, rakentuvatko sanat jokelteluissa käytetyistä tavurakenteista ja äänteistä
●Kaksosparin lasten kehityksen yhteneväisyyksiä ja eroja: käyttävätkö lapset samoja äänteitä, tuottavatko samoja sanoja, kuinka yksilöllistä lasten kehitys on
●Kaksos- ja yksöslasten kehityksen yhteneväisyyksiä ja eroja
● Mitkä kehitykselliset piirteet kuuluvat kaksoslasten normaaliin kielen kehitykseen ja mitkä piirteet ennakoivat kielellisiä vaikeuksia

**Tutkimukseen osallistuvien perheiden osuus**
Aineistoa kerätään 4–24 kk korjatun iän välillä. Tutkimuksen suunnittelussa on erityisesti pyritty siihen, että tutkimuksesta koituisi mahdollisimman vähän lisäkuormitusta perheille. Aineisto koostuu kotioloissa tehdyistä nauhoituksista, joiden lisäksi täytetään lomakkeet 1;0, 1;6 ja 2;0 ikäpisteissä.

Nauhoitukset
● Nauhoitukset tehdään perheiden kotona 1-2 kk välein.
● Nauhoituslaitteet tuodaan perheen kotiin aamulla ja haetaan nauhoituksen päätyttyä
●Nauhoitustilanteen aikana perhe voi toimia kotona normaalisti oman päivärytminsä mukaan ilman perheen ulkopuolisia henkilöitä.
● Aloitetaan neljän kuukauden korjatusta iästä (lasketaan lasketusta ajasta).
● Nauhoitukset päättyvät lasten ensisanoihin

Lomakkeet
●Lomakkeet postitetaan perheille kotiin palautuskuorilla varustettuina.
●Lasten motorista ja kielellistä kehitystä kartoittava lomake (yksi/lapsi) täytetään 4-12 kk iässä
●Lasten sanastoa kartoittavat lomakkeet täytetään 1;0, 1;6 ja 2;0 korjatun iän ikäpisteissä.

Tutkimukseen osallistuvien perheiden kanssa tehdään sopimus aineiston käytön oikeuksista. Tutkimukseen osallistuminen on vapaaehtoista ja tutkimuksen keskeyttäminen on mahdollista koko tutkimuksen ajan. Aineiston säilytyksessä, käsittelyssä ja tutkimustulosten raportoinnissa huolehditaan tutkittavien yksilönsuojan toteutumisesta. Tutkimuksen loputtua perheet saavat halutessaan koonnit omien lastensa kehityksellisistä profiileista.

Vastaan mielelläni lisäkysymyksiin!

Ystävällisin terveisin

Hanna Elo
Puheterapeutti, nuorempi tutkija

p. 040 1901 333, 040 5200 205
hanna.elo@uta.fi
http://kaksostenpuhe.weebly.com/

TAMPEREEN YLIOPISTO
Logopedia

| Sopimuslomake | Lomake luotu |
|---|---|
| | 4.1.2012 |



**SUOSTUMUS ÄÄNITALLENTEIDEN NAUHOITTAMISEEN JA ÄÄNITALLENTEIDEN SEKÄ LOMAKKEIDEN KÄYTTÖÖN**

Annan luvan lasteni

________________________________________ ja
(lapsen nimi)

________________________________________,
(lapsen nimi)

äänitallenteiden nauhoittamiseen, nauhoitteiden ja lomakkeilla kerättävien tietojen hyödyntämiseen sekä arkistointiin. Nauhoitteet ja tutkimuslomakkeet arkistoidaan Tampereen yliopiston logopedian koulutusohjelman tutkimusarkistoon.

Tallennetun aineiston arkistointiin ja jatkokäyttöön annan luvan seuraavin ehdoin:

Arkistointipaikka: Tampereen yliopisto, logopedia

Tallenne, jota lupa koskee: Kaksoslasten jokeltelun ja varhaisen sanaston kehitys-tutkimuksen yhteydessä kerättävä aineisto

| Arkistoitua materiaalia saa jatkossa käyttää: | **kyllä** | **ei** |
|---|---|---|
| 1. Kaksosten jokeltelun ja sanaston varhainen kehitys- tutkimuksessa, tutkimusraporteissa ja esitelmissä | ……. | ……. |
| 2. Myöhempien tieteellisten julkaisujen ja esitelmien teossa (esim. opinnäytetyöt, artikkelit) | …… | ……. |
| 3. Opetuskäyttöön pienryhmässä (läsnäolijoilla vaitiolosopimus) | ……. | ……. |

Tallenteita koskevat toiveet ja rajoitteet:________________________________________

________________________________________________________________

Tietosuoja: 

Aineistonkeruu, aineiston käsittely ja aineiston säilyttäminen toteutetaan Tampereen yliopiston tutkimuseettisten periaatteiden mukaisesti. Aineiston käyttäjiä sitoo vaitiolovelvollisuus. Tutkimuksen tiedot säilötään tunnisteellisina, jotta myöhemmät yhteydenotot tutkittaviin mahdollistuvat. Jatkoyhteydenottoja varten säilytettävät tutkittavien tunnistetiedot suojataan ja säilytetään analysoitavasta aineistosta erillään logopedian koulutusohjelman arkistossa. Tutkimusraporteissa huolehditaan yksityisyyden suojasta siten, että tutkittavia ei voi raporttien perusteella tunnistaa.

Olen tietoinen siitä, että voin halutessani peruuttaa tämän suostumuksen. Tallenteiden vastuuhenkilö on opetuskoordinaattori Anna Oksa p. 050 4211063.

Alaikäisiä koskeva arkistointilupa raukeaa hänen saavuttaessaan täysi-ikäisyyden. Lupa on voimassa _________________ asti.

Paikka ja päiväys Allekirjoitus

__________________ ________________________________________

**Vanhemman yhteystiedot:**

Nimi:____________________________________________________________________

Osoite:__________________________________________________________________

Puhelin:________________________ Sähköpostiosoite:_________________________

Tampereen yliopisto, yhteiskunta- ja kulttuuritieteiden yksikkö/logopedia, Päätalo, 33014 Tampereen yliopisto
Puhelin 050 4211063
(Lomake muokattu HY:n ja Turun yliopiston logopedian oppiaineen lomakkeesta)

Tutkimustiedote

## Suomalaisten kaksoslasten jokeltelun ja varhaisen sanaston kehitys

Kaksoslapset ovat puheen ja kielen kehityksen osalta riskiryhmä (Käypä hoito- työryhmä, 2010) ja tiedetään, että varhaisella puheen ja kielen kehityksellä on yhteyttä sekä myöhempiin lukemisen ja kirjoittamisen valmiuksiin että laajempiin oppimisvaikeuksiin. Suomessa puheterapeuteilla on käytössä yksöslapsilla standardoituja ja normitettuja tutkimusmenetelmiä. Ei kuitenkaan tiedetä, voidaanko kaksoslasten ja yksöslasten kielellistä kehitystä ylipäätään vertailla: Suomalaisten kaksoslasten puheen ja kielen kehityksestä on tutkimuksellista tietoa hyvin vähän ja tutkimus keskittyy pääosiin pienillä aineistoilla tehtyihin tapaustutkimuksiin (ks. Elo, 2010; Launonen, 1987; Räisänen, 1975; Savinainen-Makkonen 2000). Kaksostutkimus on kansainvälisesti keskittynyt erityisesti genetiikan vaikutukseen kielen kehityksessä identtisiä ja ei-identtisiä kaksosia vertaillen (esim. Dale, Bishop & Plomin, 2005) sekä kaksoslasten kielitaidon kuvaamiseen leikki- ja kouluikäisillä lapsilla (esim. Lewis & Thompson, 1992). Tutkimuksellista tietoa kaksoslasten varhaisesta jokeltelun, puheen ja kielen varhaisesta kehityksestä ei juurikaan ole saatavilla.

Väitöskirjani tavoitteena on saada uutta tietoa suomalaisten kaksoslasten ääntelyn varhaisesta kehityksestä kuvaamalla kaksosparien jokeltelun kehitystä ensisanoihin sekä tarkastelemalla jokeltelun kehityksen yhteyttä varhaisen sanaston hallintaan. Tutkimukseni pyrkii vastaamaan seuraavanlaisiin kysymyksiin: Milloin lasten jokeltelu alkaa ja millä lailla eri jokeltelutyypit ilmenevät ja kehittyvät. Tutkin myös, miten lasten ääntöpaikkojen ja -tapojen käyttö muuttuu jokeltelun aikana, millaisia tavurakenteita lapset tuottavat jokeltelun eri vaiheissa ja miten tavurakenteiden käyttö kehittyy. Sanaston ja jokeltelun yhteyttä tutkin tarkastelemalla jokeltelun alkamisajankohdan, määrän ja laadun yhteyttä lasten ensisanojen ilmenemisajankohtaan sekä lasten sanaston hallintaan 1;6 ja 2;0 korjatussa iässä Tarkastelen, ennustaako jokeltelun määrä ja laatu myöhempää sanaston hallintaa ja millaiset ovat kaksoslasten sanavarastot verrattuna yksöslasten sanavarastoon.

Aineisto kerätään lasten kotioloissa käyttäen pienikokoista kovalevytallenninta, joka sijoitetaan lapselle puettavan liivin taskuun. Aineistonkeruussa pyritään mahdollisimman pieneen perheiden kuormittavuuteen: Nauhurit ja liivit tuodaan perheiden kotiin ja haetaan nauhoituspäivän loputtua. Lisäksi perheille tulee koko kahden vuoden tutkimusaikana täytettäväksi kaksi erilaista tutkimuslomaketta, jotka postitetaan perheille kotiin palautuskuorilla varustettuina. *Ääntelyn ja motoriikan kehitys*-lomakkeella (Lyytinen, Ahonen, Eklund & Lyytinen, 2000) havainnoidaan lapsen ääntelyn ja motoriikan kehitystä 4-12 kk korjatussa iässä merkitsemällä ylös ikäkuukaudet, jolloin lapsella on havaittu uusia taitoja. *Varhaisen kommunikaation ja kielen kehityksen arviointimenetelmä*-lomakkeen (Lyytinen, 1999) avulla tarkastellaan lasten ymmärtävää ja tuottavaa sanastoa. Lomakkeet täytetään 1;0, 1;6 ja 2;0 korjatun iän ikäpisteissä.

Yhteydenotot ja lisätiedustelut

Hanna Elo
Puheterapeutti, vs. logopedian lehtori
Päätalo E215
33014 Tampereen yliopisto
Puh. (03) 3551 4086, 040 5200 205
hanna.elo@uta.fi