

Auteursrechterlijke overeenkomst

Opdat de Universiteit Hasselt uw eindverhandeling wereldwijd kan reproduceren, vertalen en distribueren is uw akkoord voor deze overeenkomst noodzakelijk. Gelieve de tijd te nemen om deze overeenkomst door te nemen, de gevraagde informatie in te vullen (en de overeenkomst te ondertekenen en af te geven).

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling met

Titel: Association between imitation recognition and imitation performance in typically developing children
Richting: master in Applied Statistics **Jaar:** 2008

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Ik ga akkoord,

EBEN, Agbor Tarhebi

Datum: 5.11.2008

Association between imitation recognition and imitation performance in typically developing children

Agbor Tarhebi Eben

promotor :

Prof. dr. Herbert THIJS

Dr. Raf MEESEN

De heer Tom JACOBS

CERTIFICATION

This is to certify that EBEN AGBOR TARHEBI has successfully completed the project titled "Association between imitation recognition and imitation performance in typically developing children". His work has been read and approved and thus has contributed to scientific knowledge.

EBEN AGBOR TARHEBI

.....

Student

Prof Dr. H. Thijs

T. Jacobs

.....

External Supervisor

.....

Internal Supervisor

R. MEESEN

.....

External Supervisor

DEDICATION

This piece of work is dedicated to my Parents Mr. and Mrs. EBEN JOHN, my brothers and sisters and above all to God almighty for His continual guidance and for providing me with the strength and health to go throw this work.

ACKNOWLEDGEMENT

I express my heartfelt gratitude to my supervisors Prof H.Thijs, R.Meesen, and T.Jacobs for their continuous guidance and proposals to better my work. I am also grateful to Elien Vanelderen and Koen Cornelis who initiated this study and made available the data set.

Special thanks to my friend Njinju Tongwa for his moral support and to a senior friend Assam Pryseley for academic orientation and encouragement in studies.

To my friends and mates of the University of Hasselt, I say thank you for always being there.

ABSTRACT

This project is about Imitation recognition and imitation performance in typically developing children aged within the age groups 12 to 17 months and 54 to 59 months. The objectives of the study were to explore the association between imitation recognition and imitation performance and how they influence one another.

In the study, we obtained measurements from 136 children on various ordinal variables. These observations were registered, by two observers independently, 30seconds within imitation recognition and 30 seconds to the end of the study. About 20% of the planned measurements 30seconds to the end of the study were not registered due to unwillingness of some children to cooperate in the exercise. Thus, a reflective analysis could not be obtained from this portion of the data.

Also, results based on weighted kappa indicate that there is satisfactory agreement between the measurements on Imitation Recognition (IR) from both observers. Consequently, only measurements from the first observer will be used in the analysis.

Methods for censored data were employed due to the time-to-event nature of the data, Kaplan Meier curves and frequency tables were use to explore the data with respect to Imitation recognition. Cox proportional hazards models were employed to investigate the association between imitation recognition and imitation. The forward, backward and stepwise variable selection techniques were used for variable reduction in the model building process. The final model was formed by a union of all variables in the various final automatic models and reducing the model based on Likelihood ratio test.

The Cox models indicate that the ability for a child to exhibit imitation recognition increases with increase in KIRaanhouden30sec and KIRsociaalkijken30sec. Also, the proportional hazards assumption seemed plausible for the data and the deviance residuals indicated that the model had an acceptable fit on the individual observations.

TABLE OF CONTENTS

1. INTRODUCTION	7
2. DATA.....	8
2.1 EXPERIMENTAL DESIGN	8
2.2 VARIABLE DESCRIPTION	8
3. STATISTICAL METHODOLOGIES.....	11
3.1 KAPPA STATISTICS	11
3.2 EXPLORATORY DATA ANALYSIS	13
3.2 STATISTICAL ANALYSIS	14
3.2.1 <i>Kaplan-Meier curve</i>	15
3.2.2 <i>Log Rank test and Gehan-Wilcoxon tests</i>	15
3.2.3 <i>Cox Proportional hazards Model</i>	16
3.2.4 <i>Automatic Variable selection Procedures</i>	17
4. RESULTS	18
4.1 EXPLORATORY DATA ANALYSIS	18
5: DISCUSSION AND CONCLUSION	27
REFERENCES.....	29
APPENDIX A	30

LIST OF TABLES

Table 1: strength of agreement	11
Table 2: Weighted kappa values at 30sec after Imitation recognition.....	12
Table 3:Weighted kappa values at 30sec to the end.	12
Table 4: Weighted kappa values at Manipulation phase	13
Table 5: EIR by KIR	18
Table 6: Quartile Estimates FOR KIR	19
Table 7: Summary statistics for KIR	20
Table 8: summary statistics for KIRverbalisatie30sec.....	21
Table 9: summary statistics for KIRfacialeexpressie30sec	22
Table 10: Summary statistics for KIRaanhouden30sec	23
Table 11: summary statistics for KIRvocalisatie30secs	24
Table 12: The PHREG procedure	24
Table 13: Final manual model selection	25
Table 14: Proportional Hazard Assumption	26

LIST OF FIGURES

Figure 1: Kaplan-Meier curve for K-observer	19
Figure 2: Kaplan-Meier curve for KIRverbalisatie30sec.....	20
Figure 3: Kaplan-Meier curve for KIRfacialeexpressie30sec	21
Figure 4: Kaplan-Meier curve for KIRaanhouden30sec	22
Figure 5: Kaplan-Meier curve for KIRvocalisatie30sec	23
Figure 6: Deviance residuals against linear predictor	26

1. INTRODUCTION

Children love to explore and discover. As they play, they are strengthening motor skills, developing social skills and discovering all that their immediate environment has to offer. In the beginning of their lives, babies explore their world through play. Touching, feeling, looking and listening are the keys to knowledge.

At the age of two to three months, infants begin to give the impression of being quite different persons. When engaged in social interaction, they appear to be more integrated. Infants by the second month focus their attention on the internal features of faces (Haith, Bergman & Moor 1997; Maurer & Salapatek 1976), spend more time in awake and alert state (Wolff 1987) and reciprocate in context of face-to-face interaction (Fogel 1993; Trevarthen 1979). There comes in imitation.

Imitation is an advanced behavior whereby an individual observes and replicates another. According to a study in the September 2005 *Archives of Pediatrics and Adolescent Medicine* what we eat and drink and smoke, children as young as two are already developing their own internal ‘scripts’ about adult social life that they will want to imitate.

Allowing the child to take the lead in his or her play is a positive form of encouragement. It is also a parent’s opportunity to observe and learn about their child’s play. Children love to have their parents join in and offer suggestions thus serves as a great opportunity for parents to teach their children. Language and communication are key skills in our world. Babies love to be spoken to. They love to observe and are quick to understand and imitate. Even if the child is too young to understand, it enhances later understanding.

Pictures and books are an opening to a child’s world of creativity and imagination. Children love to play finger games. These games encourage finger dexterity, language development, listening skills and teaches the child to follow direction. It helps in increasing attention and promotes good imitation skills and encourages the understanding of concepts such as size and shape. It allows for self expression and allows for an opportunity to have fun.

Imitation recognition is the ability of a child to recognize being imitated. The time it takes for a child to recognize being imitated varies. Some babies do not realize being imitated within the allocated time frame in the study.

Our goals for the study include;

- Investigate in a reliable manner, imitation recognition
- Explore the association between imitation recognition and imitation ability in young children

2. Data

2.1 Experimental design

Procedure for Imitation Recognition Test

It consists of 2 phases which are randomized and counterbalanced for age and sex. It is naturally expected that the older children will be more capable of Imitation Recognition than the younger children. Unfortunately information for such were not available thus this was not considered in our analysis.

The first phase is the Investigator manipulation phase. We observe how the child responds when the investigator uses the same object as the child. This process lasts 2mins in the older children and 4mins in the younger children. It consists of 12 behavioral observations on ordinal scale and 1 nominal scale data-Imitation recognition.

The second phase consist of two time frames; 30secs after the first sign of recognition and end of 30secs until the end of the phase. Each time frame consists of 12 behavioral observations on ordinal scale. The younger children aged between 12 to 17 months were assessed within 4mins and the older children between 54 to 59 months had 2mins.

It is worth taking the timing for the younger and older children into account in our analysis but unfortunately this information was not made available hence we could not distinguish the times used by the old and young children.

2.2 Variable Description

The dataset comprises of measurements obtained from 136 children based on their ability to imitate and also to recognize being imitated. The ages of the children range from 12-17 months referred to as young children while those between 54 to 59 months are referred to as older children. We naturally expect children of the older age group to (approximately 5yrs) to have abilities such as vocal ability; verbal ability as a result should perform better than the younger children (between 1yr and 2 yrs old).

The variables identified in the data set are referred to as categorical variables. This is because the measurement scales of the variables consist of a set of categories. They are further classified into Nominal and Ordinal variables.

Nominal refers to the categorical variables having unordered scales meanwhile Ordinal refers to categorical variables having ordered scales. The variables measured on the Nominal or Ordinal scales are both referred to as qualitative variables because their measurements consist of ordered

or ordered discrete categories. The following categorical variables were measured during the study.

Ordinal Variables

Subjective score imitation recognition, amount of verbalizations, amount of vocalization, amount of exaggerate behavior amount of sustained behavior, amount of Repetitive behavior, amount of facial expressions, amount of handling-looking behavior, amount of test reach behavior, amount of emotion, amount of social observing behavior, amount of turning away, and amount of imitating behavior. All these variables are classified on their respective ordered scales ranging from 0 to 3.

Nominal Variables

The following nominal variables were also recorded: Identification number of each child, Imitation recognition, live scoring and imitation recognition.

Unlike the other variables mentioned thus far, the time it takes a child to exhibit imitation recognition (*Time*) is quantitative. That is, it can take any value on the measurement scale (continuity).

Definition of variables

- IDNR: This is the identification number. Each child has a unique ID number ranging from 1 to 136. This is thus a nominal variable.
- IR: Imitation recognition for each subject measured on Nominal scale with 0 for no IR and 1 if there is IR.
- Livescoring IR: Live scoring Imitation recognition for each subject with 0 for no IR and 1 for IR.
- Subscore: subjective score imitation recognition from two observers following the study. Measurement was on an ordinal scale measurements as indicated below

0 =NO IR

1 =POOR IR

2 =MODERATE IR

3 =PRONOUNCED IR

- Subscore 2: subjective score imitation recognition after 30secs of IR with ordinal scale measurements from 0 to 3 as above.
- tijdIR: This is the time when IR is first seen. This is a quantitative variable because it has underlying continuity; that is can take any value on the measurement scale.

There were 12 behavioral observations on ordinal scale include;

- IR verbalization: amount of verbalization in the IR .It is measured as below;

0 =no verbalization

1 =a Verbalization

2 =several verbalizations

3 =frequent verbalization

- IR vocalization: amount of vocalization in the imitation recognition.
- IR overacting: amount of overacting behavior in IR scaled from 0 to 3.
- IR sustained: amount of sustained behavior in the IR.
- IR repetitive: amount of repetitive behavior in IR.
- IR facial expression: amount of facial expression in IR.
- IR handling looking: amount of handling-looking behavior in IR.
- IR test reach: amount of test reaching behavior in IR.
- IR emotion: amount of emotion in IR.
- IR social eye contact: amount of social eye contact in IR.
- IR withdrawals: amount of withdrawals in IR.

The qualitative scoring system of the Imitation recognition test evaluates the responding behavior of the child. This was done at different timings and by 2 independent individuals. These raters were represented by K and E attached to the variables.

Even though it is naturally expected for children in the older age group to exhibit imitation recognition than children in the younger age group, the information of the corresponding age group for each child is not available in the data, as well as their respective gender.

3. STATISTICAL METHODOLOGIES

3.1 Kappa Statistics

Using weighted kappa values, we can calculate an inter-rater agreement statistic to evaluate the agreement between the two classifications on ordinal scale. It tells us the proportion of times raters would agree by chance alone. Kappa does not take into account the degree of disagreement between observers and all disagreement is treated equally as total disagreement.

As a result we make use of weighted kappa. We use weighted kappa values because the categories are ordered. This enables us to assign different weights W_i to subjects from whom the raters differ by i -categories so that different levels of agreement can contribute to the value of kappa.

The **K** value can be interpreted as follows (Altman 1991):

Table 1: strength of agreement

Value of K	Strength of agreement
0.20	Poor
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Good
0.81-1.00	Very good

This implies that the higher the k-value the higher the strength of agreement between the two raters.

Table 2: Weighted kappa values at 30sec after Imitation recognition

Variable	Weighted kappa
KIRverbalisatie by EIRverbalisatie	0.5517
KIRvocalisatie by EIRvocalisatie	0.6608
KIROverdrijven by EIROverdrijven	0.4253
KIRaanhouden by EIRaanhouden	0.5919
KIRrepetitief by EIRrepetitief	0.5427
KIRfacialexpressie by EIRfacialexpressie	0.2929
KIRhandelkijk by EIRhandelkijk	0.0591
KIRtestgrijp by EIRtestgrijp	0.4301
KIRsociaalkijken by EIRsociaalkijken	0.1123
KIRafkeren by EIRafkeren	-
KIRimitatie by EIRimitatie	0.6595

From the table above, we realize that there is generally a fair agreement in the results from both observers. Our variable of interest has particularly a good strength of agreement between observers thus an analysis of the variable from one observer could be reflective of the other.

Table 3: Weighted kappa values at 30sec to the end.

To avoid undefined results, weighted kappa could not be computed in all cases(-).

Variable	Weighted kappa
KIRverbalisatieEINDE by EIRverbalisatieEINDE	-
KIRvocalisatieEINDE by EIRvocalisatieEINDE	0.7259
KIROverdrijvenEINDE by EIROverdrijvenEINDE	0.1683
KIRaanhoudenEINDE by EIRaanhoudenEINDE	-
KIRrepetitiefEINDE by EIRrepetitiefEINDE	0.6470
KIRfacialexpressieEINDE by EIRfacialexpressieEINDE	0.4310
KIRhandelkijkEINDE by EIRhandelkijkEINDE	0.3116
KIRtestgrijpEINDE by EIRtestgrijpEINDE	-
KIRmotieEINDE by EIRmotieEINDE	-
KIRsociaalkijkenEINDE by EIRsociaalkijkenEINDE	-
KIRafkerenEINDE by EIRafkerenEINDE	0.4062
KIRimitatieEINDE by EIRimitatieEINDE	-

Table 4: Weighted kappa values at Manipulation phase

Variable	Weighted kappa
KMAverbalisatie by EMAverbalisatie	-
KMAvocalisatie by EMAvocalisatie	0.7564
KMAoverdrijven by EMAoverdrijven	0.4836
KMAaanhouden by EMAaanhouden	-
KMArepititief by EMArepititief	0.4431
KMAfacialexpressie by EMAfacialexpressie	0.6380
KMAhandelkijk by EMAhandelkijk	-
KMAtestgrijp by EMAtestgrijp	-
KMAemotie by EIRMAemotie	-
KIRMA sociaalkijken by EIRMA sociaalkijken	0.1833
KIRMAafkeren by EIRMAafkeren	0.2198
KMAimitatie by EIRMAimitatie	0.4690

3.2 Exploratory Data Analysis

Exploratory data analysis was introduced by John Turkey as an approach to analyze data when there is only a low level of knowledge about its cause system as well as its contextual information. EDA aims at letting the data itself influence the process of suggesting hypothesis instead of only using it to evaluate a given hypothesis. Thus exploratory data analysis is a detective work (Turkey, John (1977), *Exploratory Data Analysis*, Addison –Wesley).

In this section we explore the data in formal tabulations and graphical displays to serve as the foundation stone for our analysis. The response or dependent variable is Imitation recognition and is influenced by several independent variables. These behavioral observations- independent variables were measured 30secs after Imitation recognition and 30secs to the end of the experiment.

Of the 136 children, in the k-observer, 97 exhibited Imitation recognition and 38 children registered no Imitation recognition with percentage 71.3% and 27.9% respectively. Child with IDNR 12 did not participate in the test accounting for the rest 0.8%. This should have resulted from aggressive behavior, crying and children not willing to cooperate. In the E-observer, 102 children exhibited IR while 34 registered no Imitation recognition accounting for 75% and 25% respectively.

Imitation recognition is measured on a nominal scale with 0 representing no Imitation recognition and 1 representing Imitation recognition. The time it takes for imitation recognition is equally very important as some children imitate faster than others but they are all categorized

as 1. In cases of no imitation recognition, it could possibly be due to insufficient timing. This means if time is unlimited then imitation recognition should be possible in all cases.

Just as in many biomedical applications, the primary end point of interest is the time it takes for a certain event to occur in which case it will be the time it takes for Imitation recognition to occur. The data collected is over a finite period of time-2mins for the older children and 4mins for the younger children. Therefore the time to Imitation recognition may not be observed for all individuals in our study population. The amount of follow up for the various children varies from subject to subject.

It should be noted that we are encountered with time-to-event data. This is a consequence of the fact that in addition to exhibiting imitation recognition, the time it takes a child to exhibit imitation recognition is also important. Thus, we make use of methods for censored data or survival data, in particular the Cox proportional hazards model.

3.2 Statistical Analysis

Survival analysis is a body of methods used in analyzing time to event data or failure time data. In this case the time to event is the time to Imitation recognition.

In routine data analysis, we may first present some summary statistics such as mean, standard error for the mean. In analyzing survival data however because of possible censoring, the summary statistics may not have the desired statistical properties such as unbiasedness. For example the sample mean is no longer an unbiased estimate of the estimator of the population mean.

To investigate Imitation recognition in a reliable manner, we use a powerful and non-parametric method for estimating the survival function from a data set possibly containing censored observations, the Kaplan-Meier estimator and curves. It is advantageous as it does not assume a distribution for the data and takes into account the censored data. A survival curve plots the survivor functions $S(t)$ versus time (t) .

The Kaplan-Meier survivor function is an attempt to recover the survivor function that would have been observed if there was no censoring. The survival curve is drawn as a step function. This allows us to estimate the median survival time. The median survival time is the time at which half the subjects have reached the event of interest in which case is Imitation recognition.

The median Imitation recognition time indicates for each age group where majority of the children are captured with 95% confidence. It is unbiased unlike the mean in survival analysis. The median is a robust estimate.

To explore the association between imitation recognition and imitation performance, we explore the association between imitation recognition and each covariate. We use the Cox proportional

hazards model to model covariates and response. It is reliable with well known assumptions. It assumes proportionality of all predictors. In selecting the best fit model between covariates and response we use the automatic selection process. We calculate the hazard ratio. An exponent of which tells us the association. The hazard rate is the unobserved rate at which events occur.

The proportional hazards model proposed by Cox (1972) has been used primarily in medical testing analysis to model the effect of secondary variables on survival or time-to-event responses. It assumes that changing the explanatory variables has the effect of multiplying the hazard rate by a constant.

3.2.1 Kaplan-Meier curve

A plot of the Kaplan-Meier estimate of the survival function is a series of horizontal steps of declining magnitude which when large enough sample is taken approaches the true survival function for that population. On the curve, small vertical tick-marks indicate losses-where children have been censored.

Kaplan-Meier curve is a non parametric approach used for estimating survival distribution. If the estimated survival functions for two groups of survival data are approximately parallel (do not cross) the assumption of proportional hazards may be justified. Formal test are used to check the assumption of proportional hazards.

3.2.2 Log Rank test and Gehan-Wilcoxon tests

The log rank test statistic compares estimates of the hazard functions of the two groups at each observed event time. It is constructed by computing the observed and expected number of events in one of the groups at each observed event time and then adding these to obtain an overall summary across all time points where there is an event.

Let $j = 1, \dots, J$ be the distinct times of observed events in either group. For each time j , let N_{1j} and N_{2j} be the number of subjects "at risk" (have not yet had an event or been censored) at the start of period j in the groups respectively. Let $N_j = N_{1j} + N_{2j}$. Let O_{1j} and O_{2j} be the observed number of events in the groups respectively at time j , and define $O_j = O_{1j} + O_{2j}$.

Given that O_j events happened across both groups at time j , under the null hypothesis O_{1j} has the hypergeometric distribution with parameters N_j , N_{1j} , and O_j . This distribution has expected value

$$E_j = O_j \frac{N_{1j}}{N_j}$$

and variance

$$V_j = \frac{O_j(N_{1j}/N_j)(1 - N_{1j}/N_j)(N_j - O_j)}{N_j - 1}.$$

The logrank statistic compares each O_{1j} to its expectation E_j under the null hypothesis and is defined as

$$Z = \frac{\sum_{j=1}^J (O_{1j} - E_j)}{\sqrt{\sum_{j=1}^J V_j}}.$$

The log rank test is equivalent to the **Mantel-Haenszel** method. Actually the two differ a bit in how they deal with multiple deaths at exactly the same time point.

The log-rank test is more standard. It is the more powerful of the two tests if the assumption of proportional hazards is true. Proportional hazards means that the ratio of hazard functions (deaths per time) is the same at all time points. One example of proportional hazards would be if the control group died at twice the rate as treated group at all time points.

The Gehan-Wilcoxon method gives more weight to deaths at early time points, which makes lots of sense. But the results can be misleading when a large fraction of patients are censored at early time points. In contrast, the log-rank test gives equal weight to all time points. The Gehan-Wilcoxon test does not require a consistent hazard ratio, but does require that one group consistently have a higher risk than the other.

3.2.3 Cox Proportional hazards Model

Proportional hazards models are a sub-class of survival models in statistics. We consider survival models to consist of two parts: the underlying hazard function, describing how hazard (risk) changes over time and the effect parameters, describing how hazard relates to other factors - such as the choice of treatment, in a typical medical example. The *proportional hazards assumption* is the assumption that effect parameters multiply hazard: for example, if taking drug X halves your hazard at time 0, it also halves your hazard at time 1, or time 0.5, or time t for any value of t . The effect parameter(s) estimated by any proportional hazards model can be reported as hazard ratios.

Sir David Cox observed that if the proportional hazards assumption holds (or, is assumed to hold) then it is possible to estimate the effect parameter(s) without any consideration of the hazard function. This approach to survival data is called application of the *Cox proportional hazards model*, sometimes abbreviated to *Cox model* or to *proportional hazards model*.

The generic term *parametric proportional hazards models* can be used to describe proportional hazards models in which the hazard function is specified. The Cox proportional hazards model is sometimes called a *semi-parametric* model by contrast.

Some authors (e.g. Bender, Augustin and Blettner, Statistics in Medicine 2005) use the term *Cox proportional hazards model* even when specifying the underlying hazard function, to acknowledge the debt of the entire field to David Cox. The term *Cox regression model* (omitting *proportional hazards*) is sometimes used to describe the extension of the Cox model to include

time-dependent factors. However, this usage is potentially ambiguous since the Cox proportional hazards model can itself be described as a regression model.

3.2.4 Automatic Variable selection Procedures

The aim is to identify subset of variables upon which the hazard function depends. Because there is a pool of p potential explanatory variables, automatic routine of variable selection were also considered.

Forward selection: In building the model, variables with the smaller p value are added sequentially. At each stage, the variable included in the model is the one that gives the largest decrease in the value of $-2\loglik$ on its inclusion. If there are no additional variables meet the 0.25 sign level of entry into the model then variables cannot be included again into the model. Amongst all the variables for addition to the model, the one with the *largest* effect on the criterion (AIC, LR) is added so the more the covariate is added, the more the AIC and LR decreases but none is added if AIC increases (AIC) or LR not significant (LR).

Backward selection: We first fit a model considering all possible variables. Variables with the large p - value ($p\text{-value} > 0.15$) will be excluded from the model one at a time until there are no (additional) variables that meet the 0.15 level for removal from the model. At each stage, the variable omitted is the one that increases the value of $-2\loglik$ by the smallest amount on its exclusion. Amongst the variables for elimination from the model, the one with the smallest effect on the criterion (AIC, LR), is eliminated so the more the covariate is eliminated, the more the AIC decreases and LR increases but none is eliminated if AIC increases (AIC) or LR significant (LR)

Step-wise procedure: In this method, a variable included at an earlier stage in the model building can be removed at a later stage. Thus after adding a variable to the model, the procedure then checks whether any previously included variable can now be deleted.

These procedures of covariate selection were chosen because the use of one or more of them will give rise to a subset of statistically significant covariates.

4. RESULTS

4.1 Exploratory Data Analysis

From the weighted kappa values for the variable of interest, Imitation Recognition, we observed that there is a good strength of agreement between the two raters. The simple kappa statistics for Imitation recognition from the 2 raters is 0.6595 with asymptotic standard error of 0.0735 and 95% confidence limits [0.5155; 0.8035]. The values obtained confirm good strength of agreement.

Table 5: EIR by KIR

KIR	EIR(E- imitation recognition)				
	Outcome	Summary Statistics	0	1	Total
	0	Frequency	27.00	11.00	38
	0	Percent	20.00	8.15	28.15
	0	Row Pct	71.05	28.15	
	0	Col Pct	79.41	10.89	
	1	Frequency	7.00	90.00	97
	1	Percent	5.19	66.67	71.85
	1	Row Pct	7.22	92.78	
	1	Col Pct	20.59	89.11	
	Total	0	34.00	101.0	135
		1	25.19	74.81	100

There are a total of 136 children, 1 child did not participate in the exercise and no entries were registered for this child. As a consequence our analysis does not include this child. In the K-observer, 97 children showed imitation recognition while 38 did not show imitation recognition accounting for 71.85% and 28.15%, respectively. In the E-observer, 101 children show imitation recognition while 34 children show no imitation recognition accounting for 74.81% and 25.19% respectively.

Figure 1 shows the survival curve for KIR from observer- K, obtained by the Kaplan-Meier method. The survival curve is drawn as a step function and the values range from 0 to 1. It depicts the probability of not being able to exhibit imitation recognition (henceforth referred to as “No imitation recognition”) against time. Some observations were censored as represented in the curve as a result, the downward sloping curve does not reach the zero mark of the survival axis.

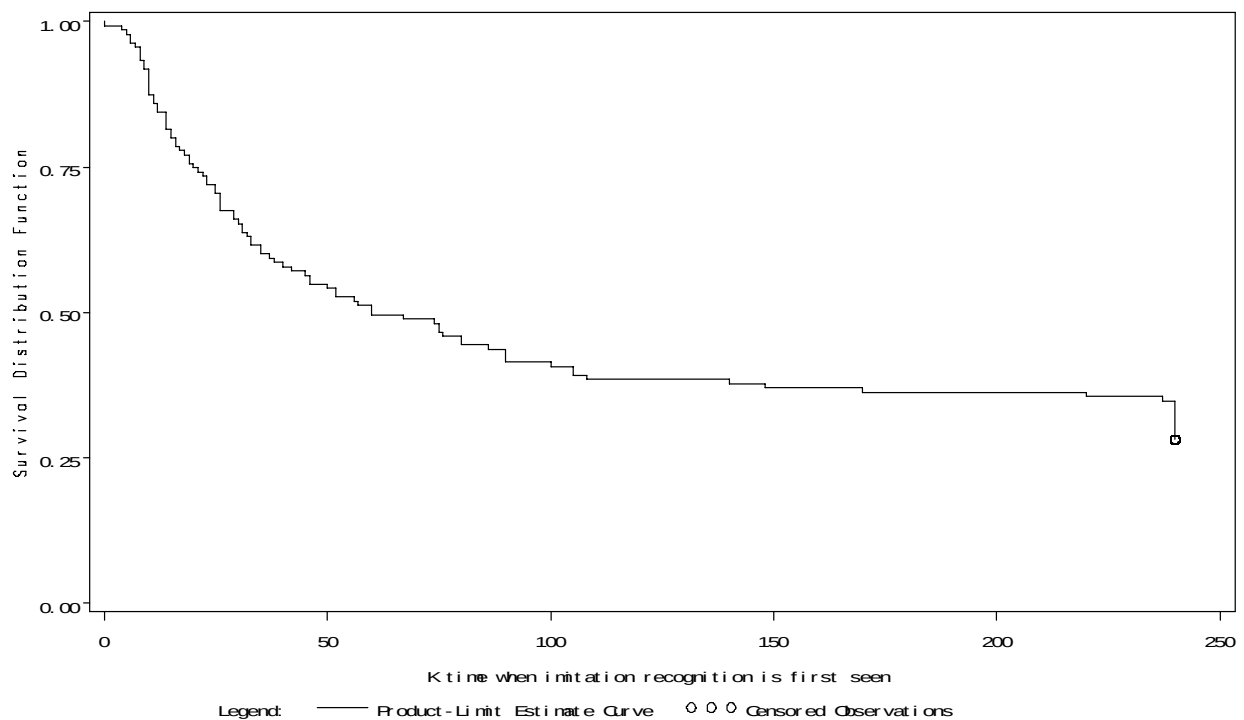


Figure 1: Kaplan-Meier curve for K-observer

The graph indicates that at time 0 no child showed Imitation recognition. Therefore the probability of “No Imitation recognition” is 1 since the study is yet to begin. Whenever there is a downward step in the curve it indicates that some children are exhibiting Imitation recognition otherwise it will indicate that there was no Imitation recognition. Censoring is observed in the curves as the curve does not drop to the zero mark on the survival axis.

Table 6: Quartile Estimates FOR KIR

Percent	Point Estimate		95% Confidence Interval	
			[Lower	Upper)
75	58.500	42.000	76.000	
50	27.500	22.000	35.000	
25	14.000	10.000	18.000	

By extrapolation, the median survival time is 60secs. It represents the time at which half the children have exhibited Imitation recognition. The 50-percentile, lower and upper quartiles for the point estimates are captured in the intervals 27.5, 22.0 and 35.0 respectively. Majority of the children exhibited imitation recognition while a few did not.

The 50th percentile has a point estimate of 27.5 indicating asymmetry. The mean value is 43.693 with a standard error of 4.813. The results obtained from KIR are similar to those of EIR. The Kaplan-Meier curve and point estimates of EIR are very similar. This confirms the good strength of agreement between the two observers. Thus results obtained from one observer could be inferred for the other.

Table 7: Summary statistics for KIR

	Number of children	Percentage
Imitation Recognition	97	71.3
No imitation Recognition	38	27.9
Non participant	1	0.8

Kaplan-Meier curves between the response and each covariate were obtained to investigate the association between each covariate and the response. Below are some of the Kaplan-Meier curves for the various variables.

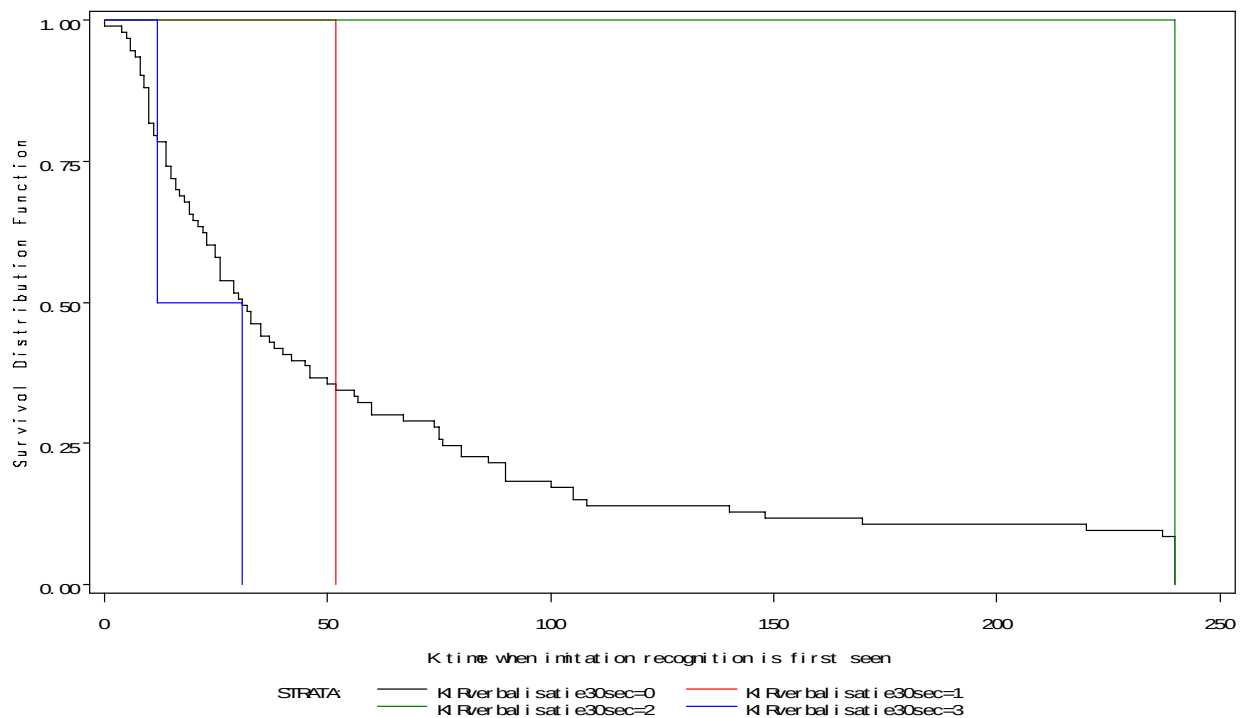


Figure 2: Kaplan-Meier curve for KIRverbalisatie30sec

In this curve there are 4 different strata indicating the strata of the children. They range from 0, 1, 2 and 3. At time 0 the experiment was yet to begin and thus no child showed imitation recognition confirmed by the horizontal line. At time when there is a fall in the curves, some

children showed imitation recognition. Thus a drop in the curve indicates that some children in the corresponding strata showed imitation recognition at that specific time point. The results obtained in strata 3, 1 and 2 are unstable and unreliable since they consist of very few children. They are thus not comparable since there is a great disparity in the number of children in the various strata. The censored children are found in stratum 0 as a result the curve for this stratum does not reach the 0 mark. It should be noted that due to the disparity in the number of children in the various strata for this variable, it was not included in the analysis.

Table 8: summary statistics for KIRverbalisatie30sec

KIRverbalisatie30sec	Event	% Event
0	84	95.3
1	1	1.2
2	1	1.2
3	2	2.3

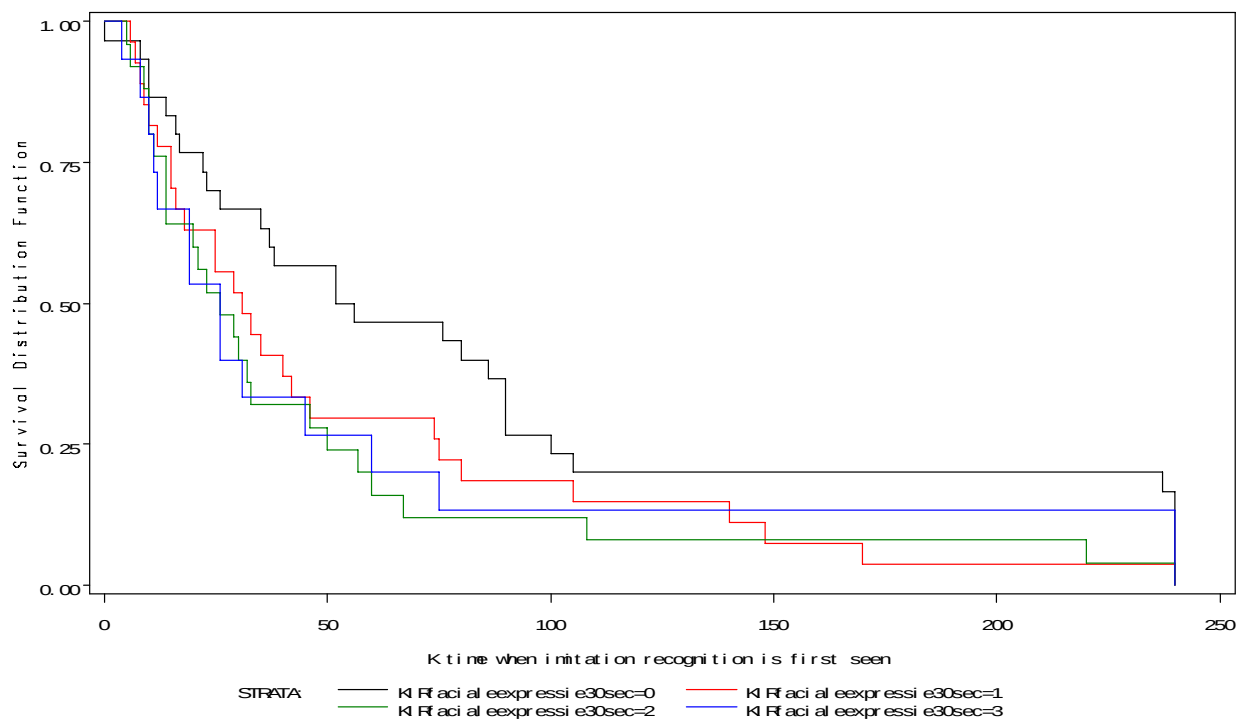


Figure 3: Kaplan-Meier curve for KIRfacialeexpressie30sec

The curves for the various strata cross each other suggesting the proportional hazard assumption may not hold. We further use the log rank test to test equality over strata because it is robust to proportionality of hazard assumption. Based on the curve, children in strata 0 appear to be slower in exhibiting imitation recognition, since the curve for this stratum appears to be consistently above those of the other strata.

Table 9: summary statistics for KIRfacialeexpressie30sec

KIRfacialeexpressie30sec	Event	% Event
0	25	28.4
1	26	29.5
2	24	27.3
3	13	14.8

The censored observations are found in strata 3 and 2 hence the curve remains horizontal and does not drop to 0. The various strata are comparable since they do not have a great disparity in their numbers of subjects. The subjects are almost evenly distributed in the various strata in this variable.

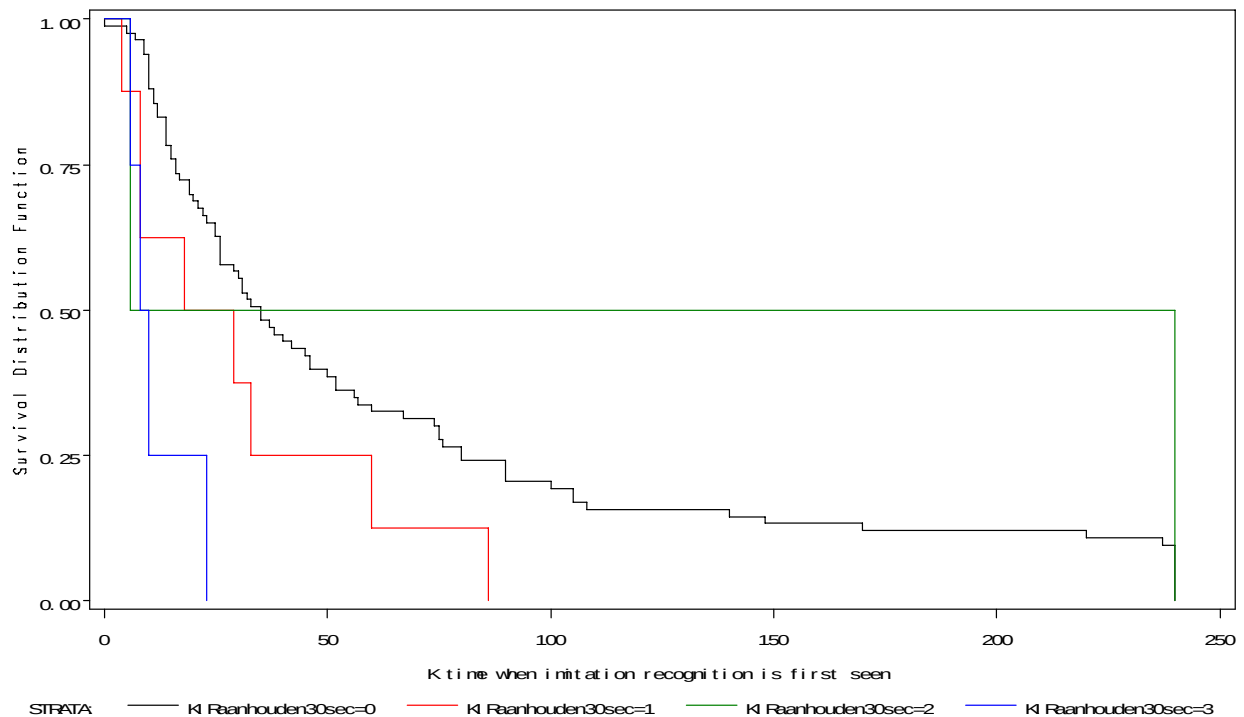


Figure 4: Kaplan-Meier curve for KIRaanhouden30sec

Stratum 0 has the censored children. The curves are not comparable due to great difference in the number of subjects in the various curves. From the curves the best children are found in strata 2, 3, 1 and 0 based on their median survival times. Strata 2 and 3 have just 1 and 4 children respectively thus the results obtained may be unstable and unreliable.

Table 10: Summary statistics for KIRaanhouden30sec

KIRaanhouden30sec	Event	% Event
0	61	69.3
1	8	9.1
2	10	11.3
3	9	10.3

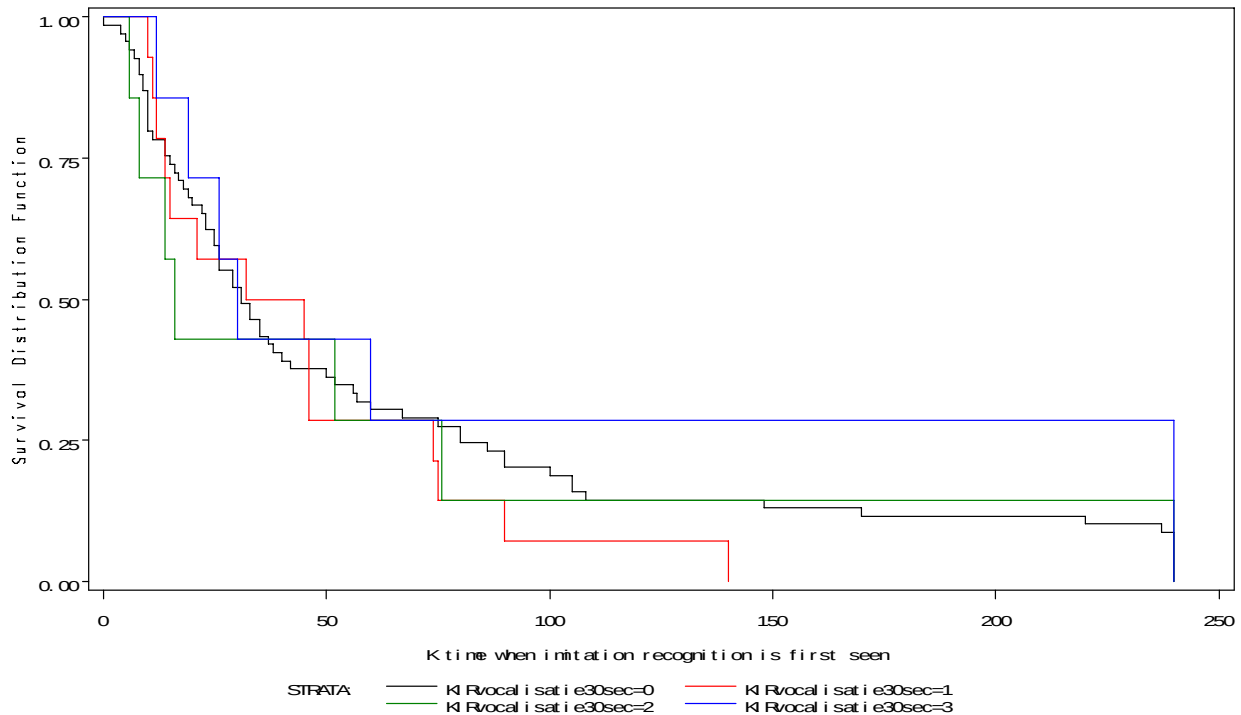


Figure 5: Kaplan-Meier curve for KIRvocalisatie30sec

The curves cut across each other indicating variable may not be included in our final model. Strata 0 and 2 do not drop to the 0 mark confirming censoring present. We cannot clearly say that children in some strata in this group are better than others because they intercross each other and depending on time one curve may lie above another.

It is worthy of note that 48 observations were censored as a result just 88 children experienced events in all the variables making up to the total of 136 children. From the results above, we realize that the variables which are significantly associated to the response-Imitation recognition are; KIRaanhouden30sec, KIRsocioalkijken30sec with p-values 0.0063 and 0.0296 respectively at 5% level of significance. The log of the parameter estimate (hazard) gives us the hazard ratio while an exponent of the hazard ratio gives us the parameter estimate.

Table 11: summary statistics for KIRvocalisatie30secs

IRvocalisatie30sec	Event	% Event
0	63	71.6
1	14	15.9
2	6	6.8
3	5	5.7

The P-values are calculated based on the chi square. The chi square helps us to visualize the contributions of the variables directly without having to look at the parameter estimate and standard errors. For example the parameter estimates for *KIRafkeren30sec* and *KIRaanhouden30secs* are 0.57929 and 0.51296. They seem close to each other but from their chi square values of 1.4261 and 7.4615 we evidently realize that *KIRaanhouden30secs* has a greater contribution.

Table 12: The PHREG procedure

Variable	DF	Parameter estimate	Standard error	Chi square	Pr ChiSq >	Hazard ratio
KIRvocalisatie30sec	1	0.09618	0.14194	0.4592	0.4980	1.101
KIRoverdrijven30sec	1	0.14037	0.23334	0.3619	0.5475	1.151
KIRaanhouden30sec	1	0.51296	0.18779	7.4615	0.0063	1.670
KIRfacialeexpressie30sec	1	0.10664	0.11795	0.8175	0.3659	1.113
KIRhandelkijk30sec	1	-0.07549	0.12996	0.3374	0.5613	0.927
KIRtestgrijp30sec	1	0.19659	0.30864	0.4057	0.5242	1.217
KIRmotie30sec	1	0	-	-	-	-
KIRsocaalkijken30sec	1	0.40399	0.18569	4.7334	0.0296	1.498
KIRafkeren30sec	1	0.57929	0.48509	1.4261	0.2324	1.785

A negative parameter estimate indicates to us that the hazard ratio is less than 1. In the case of the covariate *KIRhandelkijk30sec*, the hazard ratio of 0.927 indicates that the active level is about 92 units that of the reference level. This implies the active level has fewer events.

It is worth mentioning that scores were used; 0, 1, 2, 3 representing none poor, moderate and pronounced respectively for the various variables. The variables were treated as continuous hence have a degree of freedom equal 1. *KIRaanhouden30sec* has a hazard ratio of 1.67 indicating that the hazard of the active stratum is about 70 units more than that of the reference stratum.

KIRsocioalkijken30sec has a hazard ratio of 1.498 indicating that the hazard of the active stratum is about 50units more than that of the reference stratum. The active stratum has more events in both cases.

In fitting the model we begin with an initial full model where all the possible variables are included in the model. We further use the automatic selection procedure; forward, backward elimination and stepwise selection procedures. The final model is formed by a union of all variables in the various final automatic models and reducing the model based on Likelihood ratio test, literature and prior knowledge.

In the step-wise procedure, the variables KIRaanhouden30sec, KIRsocioalkijken30sec, entered the model in this in that order. Their p-values were 0.002, 0.0129, respectively. The variable KIRafkeren30sec was added and later removed from the model. It could not meet the 0.15 sign level of stay in the model. Model building terminates because the variable to be entered is the variable that was removed.

In the forward selection procedure the variables KIRaanhouden30sec, KIRsocioalkijken30sec, KIRfacialeexpressie30sec, and KIRrepetitief30 sec constituted the model and thereafter no other variable met the 0.25level for entry into the model.

In the case of backward elimination, the variables .Here we begin with a full model and KIRremotie30sec, KIR imitatie30sec were eliminated based on their redundancy. In the end of the backward elimination only the variables KIRaanhouden30sec, KIRsocioalkijken30sec, did not meet the 0.15level for removal from the model. The final model was formed by a union of all variables in the various final automatic models and reducing the model based on Likelihood ratio test. Below is a summary of the results obtained;

Table 13: Final manual model selection

Variable	Parameter estimate	Chi square	Pr>chi square	Hazard ratio
KIRsocioalkijken30sec	0.43249	7.5700	0.0059	1.541
KIRaanhouden30sec	0.50930	9.8912	0.0017	1.664

For KIRsocioalkijken30sec, a hazard ratio of 1.541 indicates that the chance of a child showing imitation recognition increases by 54% for an increase of 1 in the KIRsocioalkijken30sec score. Similarly the chance of a child showing imitation recognition increases by 67% for an increase of 1 in the KIRaanhouden score.

Proportional Hazard Assumption

Table 14: Proportional Hazard Assumption

Variable	P Value
KIRsocioalkijken30sec	0.8940
KIRaanhouden30sec	0.7890

The proportional hazard assumption is a key aspect of the models used in this project. It is therefore apparent that evidence supporting or against this assumption need to be verified. Table 15 holds the results obtained from verifying this assumption. P values from the tests indicate that there is evidence of the plausibility of the proportional hazard assumption for both covariates in the model. Therefore, we can interpret the results based on the proportional hazards assumption.

Checking Goodness of Fit on the final model

Deviance residuals were used to investigate the fit of the model. The deviance residuals are randomly scattered around zero (See Figure 6). In conclusion, there is no indication of a lack of fit of the model to individual observations.

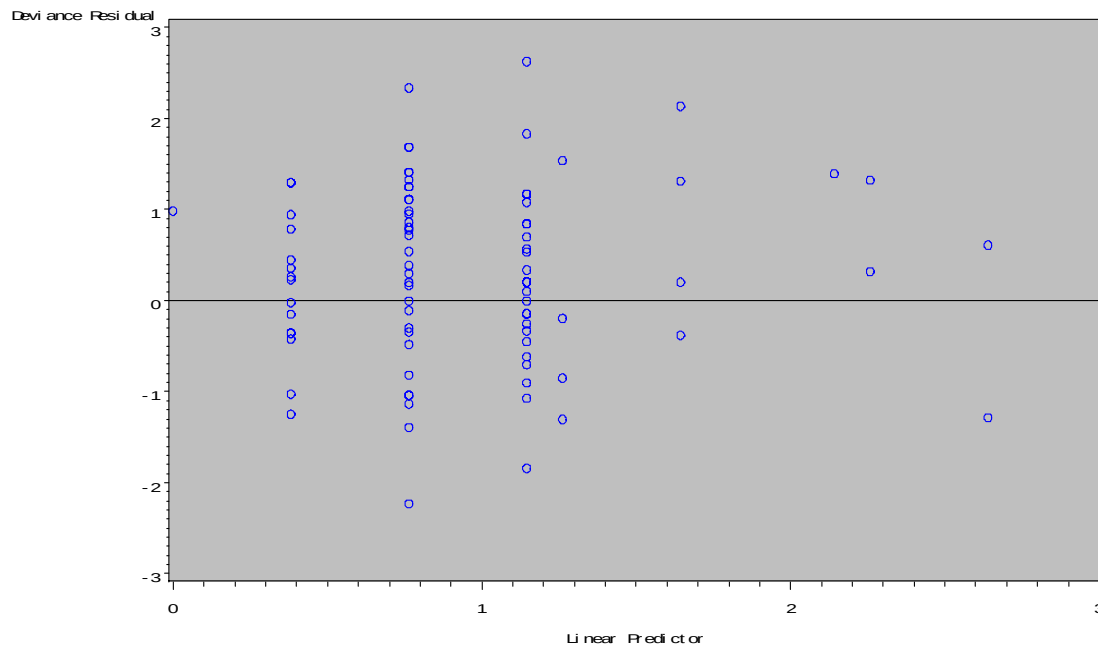


Figure 6: Deviance residuals against linear predictor

5: DISCUSSION AND CONCLUSION

Using the weighted kappa statistics, measurements on imitation recognition from both observers showed high agreement with a *K*-value of 0.6595. Due to the high inter-rater agreement, only measurements from one observer were used in the analysis.

136 Children participated in the study and were selected from two age groups, 12 -17 months and 54-59 months, which were given 2mins and 4mins respectively in the experiment. Although the study was randomized for age and sex, these variables were not available for the analysis. We naturally expect the older children to show imitation recognition faster but unfortunately our analysis could not investigate this due to the lack of information on the age groups in the data. Thus, in such a study it is imperative to obtain the ages of the individual children and the distinct timings of the different age groups' in order to get a reflective analysis of our investigation. If the age effect was taken into consideration, we would expect the children in the different age groups to be in different distinct strata and in fitting the model we fit a stratified model with different baseline hazards. This will lead us to a more representative model for our setting.

Based on the analysis above, we can conveniently say that the variables KIRaanhouden30sec and KIRsocioalkijken30sec significantly increase the ability of the children to exhibit imitation recognition. In all methods of model building used, they were predominantly significant with hazard ratios; 1.541, 1.664 in the final model. For KIRsocioalkijken30sec, a hazard ratio of 1.541 indicates that the chance of a child showing imitation recognition increases by 54% for an increase of 1 in the KIRsocioalkijken30sec score. Similarly the chance of a child showing imitation recognition increases by 67% for an increase of 1 in the KIRaanhouden score.

For a child to show imitation recognition, he must be very social. He must interact with his society and be able to respond by verbalization or by action. Verbalization happens to be one of the key skills of communication in our world of today .Some children may have the potentials of exhibiting imitation recognition but are shy so end up turning away. Thus such a variable may practically not contribute to Imitation recognition.

For the children who are bold enough to show Imitation recognition they would probably persist in the action since it opens his world of creativity and imagination. Such children are more socially integrated. They therefore develop social skills.

Finally, considering the fact that the proportional assumption seems plausible and the deviance residuals indicate a satisfactory fit of the model on the individual observations, we conclude that KIRaanhouden30sec and KIRsociaalkijken30sec are vital aspects required for exhibiting imitation recognition.

References

1. Archives of pediatrics and adolescent medicine-September 2005.
2. Report on sensitivity to social contingencies between 1 and 3 months of age Striano, 1, 3, 4. Anne Henning 1 and Daniel Stahl 2
3. Journal of the experimental analysis of behavior. Do infants show generalized imitation of gestures? By Pauline J. Horne and Mihela Erjavec University of wales Bangor Turkey, John (1977), *Exploratory Data Analysis*, Addison –Wesley. Altman DG (1991) Practical statistics for medical research. London: Chapman and Hall.
4. Fleiss JL (1981) Statistical methods for rates and proportions, 2nd Ed. New York: John Wiley & Sons.
5. Allison, P. D. (1995). *Survival Analysis Using the SAS System: A Practical Guide*. Cary, NC: SAS Institute Inc.
6. Altman, D. G. and Royston, P. (2000). What do we mean by validating a prognostic model? *Statistics in Medicine*, **19**, 453-473.
7. Derksen, S. & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noisy variables. *British Journal of Mathematical and Statistical Psychology*, **45**, 265-282.
8. Efron, B. (1988). Logistic regression, survival analysis and the Kaplan-Meier curve. *Journal of the American Statistical Association*, **83**, 414 - 425.
9. Harrell, F.E., Lee, K.L., and Mark, D.B. (1996). Multivariate prognostic models: issues in developing models, evaluating assumptions and accuracy, and measuring and reducing errors. *Statistics in Medicine*, **15**, 361-387.
10. Harrell, F. E. (2001). *Regression modeling strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag New York Inc.
11. Heinze, G. and Schemper, M. (2003). Comparing the importance of prognostic factors in Cox and logistic regression using SAS. *Computer Methods and Programs in Biomedicine*, **71**, 155-163.
12. Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression*, 1st edition. New York: John Wiley & Sons, Inc.
13. Hosmer, D. W. & Lemeshow, S. (1999). *Applied Survival Analysis*, New York:

APPENDIX A

Programmes used;

SAS 9.1 (English) , S-PLUS 6.1, SPSS 16.0

Forward

- 1 KIRaanhouden30sec
- 2 KIRsociaalkijken30sec
- 3 KIRfacialeexpressie30sec
- 4 KIRreplititief30sec

StepWise

- 1 KIRaanhouden30sec
- 2 KIRsociaalkijken30sec

Backward

- 1 KIRaanhouden30sec
- 2 KIRsociaalkijken30sec

Initial Combined Model

- 1 KIRaanhouden30sec
- 2 KIRsociaalkijken30sec
- 3 KIRfacialeexpressie30sec
- 4 KIRreplititief30sec

Final Combined Model

- 1 KIRaanhouden30sec
- 2 KIRsociaalkijken30sec

*/

```
data kappa.eben;
set 'F:\eben';
run;
proc contents;
run;
run;
proc freq data= kappa.eben;
tables KIRverbalisatie30sec * EIRverbalisatie30sec/agree all ;
*exact agree KAPPA WTKAP ;
run;
tables KIRvocalisatie30sec * EIRvocalisatie30sec/agree all ;
*exact agree KAPPA WTKAP ;
run;
tables KIROverdrijven30sec * EIROverdrijven30sec/agree all ;
*exact agree KAPPA WTKAP ;
run;
```



```

tables KIRaanhouden30sec * EIRaanhouden30sec/agree all ;
*exact agree KAPPA WTKAP ;
run;
tables KIRreplitief30sec * EIRreplitief30sec/agree all ;
*exact agree KAPPA WTKAP ;
run;
tables KIRfacialeexpressie30sec * EIRfacialeexpressie30sec/agree all ;
*exact agree KAPPA WTKAP ;
run;
tables KIRhandelkijk30sec * EIRhandelkijk30sec/agree all ;
*exact agree KAPPA WTKAP ;

libname kappa 'F:';

data kappa.eben;
set 'F:\eben';
run;
proc contents;
run;
proc freq data= kappa.eben;
tables Ksubjscore * Ksubjscore2/agree all ;
*exact agree KAPPA WTKAP ;
run;
proc lifetest plots = (s);
time KtijdIR*KIR(0);
run;
proc lifetest plots = (s);
time KtijdIR*KIR(0);
strata KIRverbalisatie30sec;
run;
proc lifetest plots = (s);
time KtijdIR*KIR(0);
strata KIRvocalisatie30sec;
run;
proc lifetest plots = (s);
time KtijdIR*KIR(0);
strata KIROverdrijven30sec;
run;
proc lifetest plots = (s);
time KtijdIR*KIR(0);
strata KIRaanhouden30sec;
run;
proc lifetest plots = (s);
time KtijdIR*KIR(0);
strata KIRfacialeexpressie30sec;
run;
proc lifetest plots = (s);
time KtijdIR*KIR(0);
strata KIRhandelkijk30sec;

```

```

run;
proc lifetest plots = (s);
time KtjdIR*KIR(0);
strata KIRtestgrijp30sec;
run;

/*
                                Fitting Proportional Hazard Models
*/

LIBNAME kappa 'F:';

DATA kappa.eben; SET 'F:\leben'; RUN;

* Initial/Full Model;
PROC PHREG DATA=kappa.eben;
    MODEL KtjdIR*KIR(0) = KIRverbalisatie30sec KIRvocalisatie30sec KIROverdrijven30sec
    KIRreptitief30sec KIRfacialeexpressie30sec KIRhandelkijk30sec KIRtestgrijp30sec KIRremotie30sec
    KIRSociaalkijken30sec KIRafkeren30sec KIRimitatie30sec KIRaanhouden30sec;
RUN;

* Automatic Model Selection;
PROC PHREG DATA=kappa.eben;
    MODEL KtjdIR*KIR(0) = KIRverbalisatie30sec KIRvocalisatie30sec KIROverdrijven30sec
    KIRreptitief30sec KIRfacialeexpressie30sec KIRhandelkijk30sec KIRtestgrijp30sec KIRremotie30sec
    KIRSociaalkijken30sec KIRafkeren30sec KIRimitatie30sec KIRaanhouden30sec
    /SELECTION=Stepwise SLENTRY=0.25 SLSTAY=0.15 DETAILS;
RUN;

PROC PHREG DATA=kappa.eben;
    MODEL KtjdIR*KIR(0) = KIRverbalisatie30sec KIRvocalisatie30sec KIROverdrijven30sec
    KIRreptitief30sec KIRfacialeexpressie30sec KIRhandelkijk30sec KIRtestgrijp30sec KIRremotie30sec
    KIRSociaalkijken30sec KIRafkeren30sec KIRimitatie30sec KIRaanhouden30sec
    /SELECTION=Forward SLENTRY=0.25 SLSTAY=0.15 DETAILS;
RUN;

PROC PHREG DATA=kappa.eben;
    MODEL KtjdIR*KIR(0) = KIRverbalisatie30sec KIRvocalisatie30sec KIROverdrijven30sec
    KIRreptitief30sec KIRfacialeexpressie30sec KIRhandelkijk30sec KIRtestgrijp30sec KIRremotie30sec
    KIRSociaalkijken30sec KIRafkeren30sec KIRimitatie30sec KIRaanhouden30sec
    /SELECTION=Backward SLENTRY=0.25 SLSTAY=0.15 DETAILS;
RUN;

* Manual Model Selection;
* Form a union of all variables in the various final automatic models;
* Reduce this model based on LRT and literature/prior knowledge;
PROC PHREG DATA=kappa.eben;
    MODEL KtjdIR*KIR(0) = KIRverbalisatie30sec KIRSociaalkijken30sec KIRaanhouden30sec;
RUN;

```