

Predictive value of Ages & Stages Questionnaires for cognitive performance during early child education

Validez del Ages & Stages questionnaires para predecir el desempeño cognitivo en los primeros años de educación escolar

Luisa Schonhaut B.^a, Marcela Pérez R.^a, Ana María Castilla F.^b, Sonia Castro M.^{a,b}, Patricia Salinas A.^a, Iván Armijo R.^c

^aServicio de Pediatría, Clínica Alemana, Facultad de Medicina, Universidad del Desarrollo, Santiago, Chile

^b"La Casa Azul", Centro Integral de Atención Psiquiátrica y Psicológica, Santiago, Chile

^cFacultad de Psicología, Universidad Gabriela Mistral, Santiago, Chile

Received: 20-11-2015; Accepted: 18-8-2016

Abstract

Introduction: The Ages and Stages questionnaires (ASQ) has been recently validated in our country for developmental screening. The objective of this study is evaluate the validity of ASQ to predict low cognitive performance in the early years of schooling. **Patients and method:** Diagnostic test studies conducted on a sample of children of medium-high socioeconomic level were evaluated using ASQ at least once at 8, 18 and/or 30 months old, and later, between 6 and 9 years old, reevaluated using the Wechsler Intelligence Scale for Children-third edition (WISC-III). Each ASQ evaluation was recorded independently. WISC-III was standardized, considering underperformance when the total score was under -1 standard deviation. **Results:** 123 children, corresponding to 174 ASQ assessments (42 of them were 8 months old, 55 were 18 months and 77 were 30 months of age) were included. An area under the ROC curve of 80.7% was obtained, showing higher values at 8 months (98.0%) compared to 18 and 30 months old (78.1 and 79.3%, respectively). Considering different ASQ scoring criteria, a low sensitivity (27.8 to 50.0%), but a high specificity (78.8 to 96.2%) were obtained; the positive predictive value ranged between 21 and 46%, while the negative value was 92.0-93.2%. **Conclusion:** ASQ has low sensitivity but excellent specificity to predict a low cognitive performance during the first years of schooling, being a good alternative to monitor psychomotor development in children who attend the private sector healthcare in our country.

Keywords:

Psychomotor development; Screening; Ages and Stages questionnaires; Learning; Cognitive development.

Introduction

Identifying psychomotor development disorders is a fundamental component of health supervision for children to receive timely treatments. Given that the

clinical impression is rather ambiguous, in particular when assessing children with no identifiable risk factors^{1,2}, the American Society of Pediatrics recommends a psychomotor development screening through standardized tests at 9, 18 and/or 30 months old³, while a

Correspondence:
Luisa Schonhaut B.
lschonhaut@alemana.cl

similar practice is observed in the Primary Health Care of our country⁴.

There are different systems for measuring a psychomotor assessment; some must be applied by a trained professional, with the drawbacks of being a time-consuming task for the professional and a different experience for the child, while others can be implemented by the parents or caregivers in the natural environment of the child³. The advantages of the latter are low cost, smooth implementation and parents' involvement⁵⁻⁷.

One of the reporting method is the Ages and Stages Questionnaires (ASQ), which was developed at the University of Oregon in the 1980s, and subsequently updated and validated in many countries^{8,9}. Over the last decade, it has gained widespread popularity in the US¹⁰. ASQ consist of a series of questionnaires targeting different age groups, which screens 5 domains, communication, gross motor, fine motor, problem solving and personal-social.

Based on the original validation, a child is at risk of developmental problem, and referred to a more specialized assessment, in case of a performance below -2 standard deviations (SD) in at least one of the questionnaires' segments⁸. However, there is debate regarding this in recent research because with a performance below the cut off in two or more domains, the ASQ specificity increases¹¹⁻¹³, while efficiency overall improves when the total score is considered^{13,14}. Independent from the approach, it has been observed that when parents conduct the assessment, the search for possible development disorders significantly increases^{2,15}, anticipating up to 30% referral increase among evaluated children^{8,16}. The current evidence is unclear as most of the research corresponds to concurrent studies; therefore, there is a need to design prospective research that helps to define a deficit criterion with a real clinical impact.

In our country, ASQ was validated in a broad sampled community¹⁷. In a sub-sample of medium-high socioeconomic level children with a seemingly typical development, the concurrent validity of ASQ was assessed with sensitivity and specificity values of 73% and 81% respectively¹⁸. A positive screening was observed in 14% of the children¹⁹; however, the future development of these children is unknown.

The objective of this study is evaluate the validity of ASQ to predict low cognitive performance in the early years of schooling in a sample of children of medium-high socioeconomic level, and also, to compare the diagnosis accuracy of many ASQ criteria as predictors of below average cognitive performance.

Patients and Method

This diagnostic test study was conducted on a sample of children of medium-high socioeconomic level

who were evaluated using ASQ at least once (at 8, 18 and/or 30 months old), between 2008 and 2011¹⁸, and later, between 6 and 9 years old, re-evaluated using the Wechsler Intelligence Scale for Children-third edition (WISC-III). Each ASQ evaluation was recorded independently.

Inclusion Criteria

Previously evaluated using ASQ at 8, 18 and/or 30 months old, without history of neurological, congenital or metabolic disease, or with a non-fluent Spanish speaker caregiver. In case of children born at less than 37 weeks gestational age, their 8 and 18 month-old ages were corrected, but at 30 months old, their chronological age was considered.

Exclusion Criteria

History of a breakthrough condition which may affect the development, such as meningitis, CNS tumor disease, vascular accident, severe head injury, diagnosed after the ASQ. Children that were living overseas or were 9 years old at the time of evaluation were also excluded.

Children with serious developmental disorders (who did not have a diagnosis at the time of ASQ), who were not in school attendance and/or able to participate in the evaluation, were not examined by WISC-III, but considered underperformance in the final analysis.

The parents of the children were contacted by phone and agreed to participate through an informed consent. Subsequently, they filled in a form regarding bio-demographic, neonatal and academic antecedents, plus possible developmental and cognitive-related diagnosis and interventions of the children. All children had at least one parent that attended college and belonged to families in the fifth income quintile.

Instruments

*Ages and Stages questionnaire*⁸

A series of 21 questionnaires for children aged 2 months-5 years. Each questionnaire screens 5 domains, 6 questions each, evaluating different aspects of the psychomotor development, such as: comunicación, gross motor, fine motor, problem solving and personal-social. They are multiple choice questions with answers «yes» (10 points), «occasionally» (5 points) o «not yet» (0 point); the scores of each domain are added for a possible total of 60 points per area. At the end, a separate section of 7 open questions to explore possible parents' concerns can be found; however, they were not considered in this analysis. The parents filled up the ASQ forms based on the age group. The questionnaires were validated in our country by Schonhaut and Armijo (2013 y 2015)^{17,18}.

Wechsler Intelligence Scale for Children-third edition²⁰

It is an individually administered clinical instrument to evaluate the cognitive performance of children between the ages of 6 and 16 and 11 months, using 13 subtests, 6 verbal scale tests and 7 performance scales. Children performance is summarized in 3 composite scores, verbal, performance and total IQ coefficients. This Scale was standardized and adapted in Chile by Ramírez and Rozas in 2007²¹. Only the total IQ coefficient was considered in this analysis.

WISC-III was implemented by a group of ASQ result-blind trained psychologists. The evaluations were performed either at the doctor's office or at the patients' home, and the scores were given after a consensual review by the participating professionals. In the end, the parents received a report of the children performance.

Definition of psychomotor development delay risk according to Ages and Stages questionnaires

Total ASQ scores were standardized and transformed into a Z-score in each age group to be analyzed as a whole. Various delay criteria were analyzed: performance < -2 standard deviation (SD) at least in one developmental area; performance < -2 standard deviation (SD) in 2 or more areas; total score < -2 DS.

Definition of low performance according to the Wechsler Intelligence Scale

The WISC-III score obtained standardized for the simple under study and a low performance was defined as the total score was 1 SD below average.

Statistical Analysis

The normality of the distribution of WISC-III and ASQ scores was analyzed using the Shapiro-Wilk test. WISC-III standardized scores, separated in two groups: underperformance and normal performance were used as development consolidation indicators. ROC curves were created to determine ASQ scores according to age group, which improve diagnosis accuracy in low performance cases.

Although there is no absolute standard to interpret the area under the curve (AUC), it is widely accepted that an AUC of 0.5 is a test without predictive value; 0.6 and 0.7 values are considered acceptable; 0.7 to 0.9 as good and higher than 0.9 as excellent²². DeLong test was used to compare the relative precision obtained among different age groups when identifying cases of low performance, comparing the AUC of all ROC curves obtained²³.

ASQ psychometric properties were analyzed (sensitivity, specificity and predictive values) using various developmental delay risk criteria previously defined.

Analyses were performed using R Statistical²⁴, specifically with the implementation of the PROC module²⁵.

This research was approved by the Faculty of Medicine, Clínica Alemana-Universidad del Desarrollo Ethics Committee.

Results

123 children with history of at least one previous ASQ assessment were included. The WISC test was used on 121 children, and 2 children with diagnosis of developmental disorders were not evaluated, but considered in the final analysis, therefore, a total of 174 ASQ evaluations were analyzed. Among these children, 73 were full term babies, 34 had been born preterm between 32 and 36 weeks and 16 had been born before 32 weeks of gestational age (table 1).

Of the total number of ASQ evaluations, 42 were performed at 8 month old, 55 at 18 months old and 77 at 30 months old. 7 children had 3 evaluations, 37 had 2 and the rest had been evaluated once (table 1).

The total score obtained using the WISC-III scale presented a normal distribution, with an average value of 114.8 and a standard deviation of 12.8 (Shapiro-Wilk statistics test for normality, $p = 0,19$). 15 children met the low performance criteria with the WISC-III scale (13 with performance below -1 SD and two children could not be evaluated), corresponding to a 12.2% of the sample.

Table 1.

	N	(%)
<i>Número de niños incluidos</i>	123	(100)
<i>Género</i>		
Masculino	70	(56,9)
Femenino	53	(43,1)
<i>Edad gestacional</i>		
RNT	73	(59,3)
PMT	34	(27,6)
PE	16	(13,0)
<i>Gemelar</i>		
Sí	33	(26,8)
No	90	(73,2)
<i>Número de evaluaciones con ASQ por niño</i>		
1	79	(64,2)
2	37	(30,0)
3	7	(5,7)
Total evaluaciones	174	(100)
<i>Edad evaluación</i>		
8 meses	42	(24,1)
18 meses	55	(31,6)
30 meses	77	(44,3)

PE: prematuros extremos (edad gestacional < 32 semanas de gestación); PMT: prematuros moderados y tardíos (32-36 semanas de edad gestacional); RNT: recién nacidos a término (37-42 semanas de gestación).

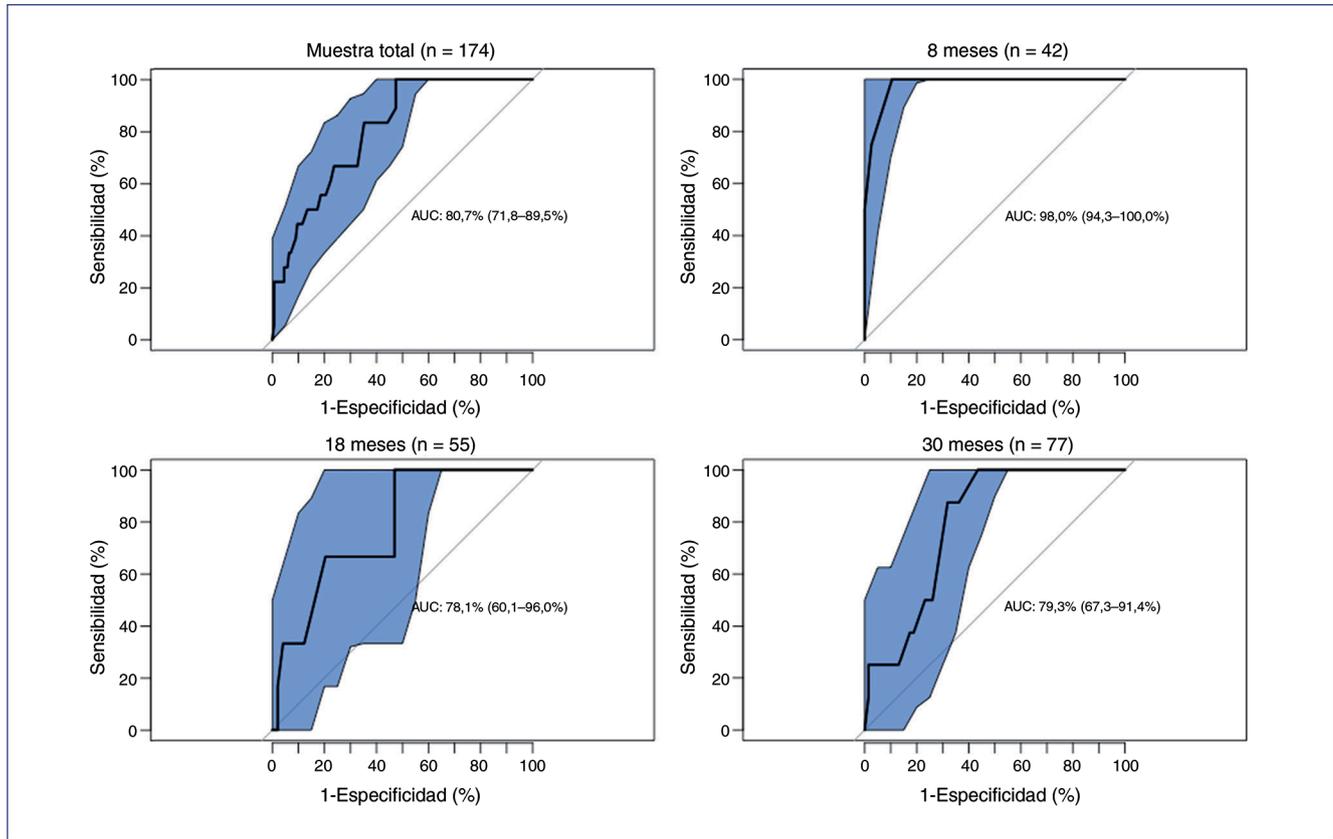


Figure 1.

Table 2. j

Criterio ASQ	Sensibilidad	Especificidad	VPP	VPN	Sobre-referencia	Infra-referencia
Análisis puntaje total $Z < -2$	33,3 (14,4-58,9)	92,9 (87,4-96,3)	35,3 (15,3-61,3)	92,4 (86,7-95,8)	64,7 (32,9-82,5)	7,6 (4,4-14,0)
Al menos un área < -2 DS	50,0 (26,8-73,2)	78,8 (71,4-84,8)	21,4 (10,8-37,2)	93,2 (87,1-96,6)	78,5 (62,8-89,2)	6,8 (3,4-12,9)
Dos o más áreas < -2 DS	27,8 (10,7-53,6)	96,2 (91,4-98,4)	45,5 (18,1-75,4)	92,0 (86,4-95,5)	54,5 (24,6-81,9)	8,0 (4,5-13,5)

DS: desviación estándar; VPN: valor predictivo negativo; VPP: valor predictivo positivo.

In ROC curve analysis, the total ASQ scores produced an area under the curve of 80.7% (CI 95%: 71.8 to 89.5). Figure 1 shows ROC curves at ages when ASQ took place with the low WISC-III performance as reference; a great precision was observed at 8 months old with an area under the ROC curve of 98.0% (CI 95%: 94.3 to 100), compared to 18 (78.1%) (CI 95%: 60.1-96.0) and 30 months (79.3%) (CI 95%: 67.3 to 91.4) where precision was overall good. A comparison of curve precision using DeLong test showed significant differences between: 8 versus 18 months ($D = -2.13$,

$gl = 58.7$, $p = 0.037$); 8 versus 30 months ($D = -2.90$; $gl = 90.0$; $p = 0.004$); but no significant differences between 18 versus 30 months were observed ($D = -0.12$; $gl = 99.16$; $p = 0.907$).

A comparison of different psychomotor development delay risk criteria showed no significant differences among ASQ psychometric properties. When considering total score < -2 SD, sensitivity values of 33.3% and specificity values of 92.9% were obtained; when at least 2 areas below the cutoff point were considered, sensitivity was 27.8% and specificity 96.2, while with at

least one area below the cutoff point, sensitivity raised to 50.0% and specificity decreased to 78.8%. Positive predictive value ranged from 21.4% to 45.5%, and over-referral percentages varied from 54.5% to 78.5%. Negative predictive values ranged from 92.0% y 93.2% in the 3 analysis criteria (table 2).

Discussion

An area under the ROC curve of 80.7% was observed in this study, a good overall discriminative capacity of the ASQ to predict a lower cognitive performance during early child education. This was reflected in adequate specificity values (78.8% to 96.2%), but low sensitivity (27.8% to 50.0%), regardless of the risk criterion used.

These results are not coincident with the studies by Halbwachs et al., who reported a high sensitivity (80%) and moderate specificity (54%) in a sample of premature infants²⁶, whereas Kerstjens et al., based on the area under the cut-off point criterion, obtained 89% and 80% sensitivity and specificity, respectively, to predict the need for special education within a year of follow-up²⁷. No other studies that show the predictive validity of ASQ to foresee lower cognitive performance in the long term have been published.

In our analysis, low positive predictive values (21.4% to 45.5%) were overall observed, with a high percentage of over-referral, 2 areas under the cut-off point criterion produced the closest rate of over-referral to the expected 30%^{8,16}. The problem of over-referral is its high economic and family costs involved²⁸, important factors to consider when establishing public health policies. Furthermore, under-referral rates were below 10%, which was acceptable⁸.

Unlike what was reported by Halbwachs using ASQ, and by Doyle using the Bayley scale^{29,30}, a tendency to improve the ASQ psychometric properties as children age increased was not observed. At 8 months, a ROC curve with an excellent discriminative capacity (AUC 98%) was obtained compared to 18 and 30 months where AUC was overall good (78.1% and 79.3%, respectively). Considering that the lower limit of confidence interval exceeded 50% in all cases, it is suggested that ASQ nonrandomly identifies children with possible developmental disorders. At 8 months old, gross motor development disorders were predominant compared to other domains³¹, which shows how intertwined the different areas of development are.

In order to obtain a correct analysis of our results, the fact that the evaluated children grew up in a culturally enriched environment and had received other evaluations and interventions, not analyzed in this study, should be considered as it could be associated with better cognitive and/or academic performance^{32,33}. The

high overall performance of children, with a low prevalence of real developmental problems, could explain the low positive predictive values and the high over-referral obtained in our study. Another limitation is the fact that it is an opportunity sample, not allowing statistical inference. Also, the fact that only cognitive development is measured without evaluating other domains of development, such as motor or social-emotional performance can be questionable, aspects not considered in the objectives of this study and coincident with other publications on the subject^{26,27}.

We concluded that, in the sample under study, ASQ has low sensitivity but excellent specificity to predict a lower cognitive performance during early child education. Considering the high percentage of over-referral, we recommend that a child be referred for a more comprehensive assessment if he or she has 2 or more areas below the cut-off point. Children who have a delay in one area, a stimulation plan and a subsequent reassessment with ASQ are suggested. Our results confirm the importance of early developmental assessment, and ASQ is a good alternative to monitor the psychomotor development of children who attend the private sector healthcare in our country.

Ethical Responsibilities

Protection of people and animals: The authors reported that the procedures followed are in accordance to the ethical standards of the responsible human experimentation committee and in agreement with the World Medical Association and the Declaration of Helsinki.

Confidentiality of the data: The authors confirmed that they have followed the protocols of their center regarding the publication of patient data.

Privacy rights and informed consent: The authors have obtained the informed consent of the patients and/or subjects referred to in the article. These documents are in the possession of the corresponding author.

Funding

Project funded by Internal Project Contests of Universidad del Desarrollo (ID 20131009161129950210) and Project Contests of Sochipe 2012.

Conflict of interests

The authors declare no conflict of interest.

Acknowledgements

We thank the psychologists Loreto González, Constanza Osorio, María Teresa Burich, Pia Margalet, Daniela Werner, Josefina Jirón and Bernardita Martínez for their support, love and dedication evaluating the children.

References

- Sheldrick RC, Merchant S, Perrin EC. Identification of developmental-behavioral problems in primary care: A systematic review. *Pediatrics*. 2011;128:356-63.
- Hix-Small H, Marks K, Squires J, Nickel R. Impact of implementing developmental screening at 12 and 24 months in a pediatric practice. *Pediatrics*. 2007;120:381-9.
- American Academy of Pediatrics. Council on children with disabilities, section on developmental behavioral pediatrics, Bright Futures Steering Committee, Medical Home Initiatives For Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics*. 2006;118:405-20.
- Normas técnicas de evaluación y estimulación del desarrollo psicomotor en el niño y la niña menor de 6 años. Ministerio de Salud. Santiago Chile 2004 [consultado 13 Nov 2015]. Disponible en: <http://www.crececontigo.gob.cl/wp-content/uploads/2013/06/Normas-tecnicas-estimulacion-y-evaluacion-desarrollo.pdf>
- Sices L, Drotar D, Keilman A, Kirchner HL, Roberts D, Stancin T. Communication about child development during well-child visits: Impact of parents' evaluation of developmental status screener with or without an informational video. *Pediatrics*. 2008;122:e1091-9.
- Schonwald A, Horan K, Huntington N. Developmental screening. Is there enough time? *Clinical Pediatrics*. 2009;48:648655.
- Schonhaut L, Armijo I, Millán T, Herreros J, Hernández K, Salgado AM, et al. Comparación entre la evaluación tradicional desarrollo psicomotor versus una prueba autoadministrada. *Rev Chil Pediatr*. 2010;81:498-505.
- Squires J, Bricker D. Ages & Stages questionnaires. Third edition. (ASQ-3): A parent-completed child-monitoring system. Stanford: Paul Brookes Publishing Company; 2009.
- Schonhaut L, Armijo I. Aplicabilidad del Ages and Stages questionnaire (ASQ) para el tamizaje del desarrollo psicomotor. *Rev Chil Pediatr*. 2014;85:12-21.
- Radecki L, Sand-Loud N, O'Connor KG, Sharp S, Olson LM. Trends in the use of standardized tools for developmental screening in early childhood: 2002-2009. *Pediatrics*. 2011;128:14-9.
- Limbos MM, Joyce DP. Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. *J Dev Behav Pediatr*. 2011;32:499-511.
- Kapci EG, Kucuker S, Uslu RI. How applicable are Ages and Stages questionnaires for use with Turkish children? *Top Early Child Spec Educ*. 2010;30:176-88.
- Flamant C, Branger B, Nguyen S, et al. Parent-Completed developmental screening in premature children: A valid tool for follow-up programs. *PLoS One*. 2011;6:4-9.
- Hornman J, Kerstjens JM, de Winter AF, Bos AF, Reijneveld SA. Validity and internal consistency of the Ages and Stages questionnaire 60-month version and the effect of three scoring methods. *Early Hum Dev*. 2013;89:1011-5.
- Guevara JP, Gerdes M, Localio R, et al. Effectiveness of developmental screening in an urban setting. *Pediatrics*. 2013;131:30-7.
- Glascoe FP, Byrne KE. The accuracy of three developmental screening tests. *J Early Interv*. 1993;17:368-79.
- Armijo I, Schonhaut L, Cordero M. Validation of the Chilean version of the Ages and Stages questionnaire (ASQ-CL) in community health settings. *Early Hum Dev*. 2015;91:671-6.
- Schonhaut L, Armijo I, Schonstedt M, Alvarez J, Cordero M. Validity of the ages and stages questionnaires in full term and preterm infants. *Pediatrics*. 2013;131:e1468-75.
- Schonhaut L, Salinas P. Evaluación del desarrollo psicomotor en las consultas pediátricas: experiencia en Clínica Alemana. *Contacto Científico*. 2012;2:48-55 [consultado 13 Nov 2015] Disponible en: <http://contactocientifico.alemana.cl/ojs/index.php/cc/article/view/46>
- Wechsler D. Test de inteligencia para niños WISC-III. Manual (traducción de Ofelia Castillo). Buenos Aires. Argentina: Paidós; 1997.
- Ramírez V, Rosas R. Estandarización del WISC-III en Chile: descripción del test, estructura factorial y consistencia interna de las escalas. *Psyche (Santiago)*. 2007;16:91-109.
- Swets A. Measuring the accuracy of diagnostic systems. *Science*. 1988;240:1285-93.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*. 1988;83:7-845.
- The R Development Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing [consultado 22 Jun 2016]. Disponible en: <http://www.R-project.org/>
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
- Halbwachs M, Muller JB, Nguyen The Tich S, et al. Usefulness of parent-completed ASQ for neurodevelopmental screening of preterm children at five years of age. *PLoS One*. 2013;8:e71925.
- Kerstjens JM, Bos AF, ten Vergert EMJ, de Meer G, Butcher PR, Reijneveld SA. Support for the global feasibility of the Ages and Stages questionnaire as developmental screener. *Early Hum Dev*. 2009;85:443-7.
- Cadman D, Chambers LW, Walter SD, Ferguson R, Johnston N, McNamee J. Evaluation of public health preschool child developmental screening: the process and outcomes of a community program. *Am J Public Health*. 1987;77:45-51.
- Halbwachs M, Muller J-B, Nguyen The Tich S, et al. Predictive Value of the parent-completed ASQ for school difficulties in preterm-born children < 35 weeks' GA at five years of age. *Neonatology*. 2014;106:311-6.
- Doyle LW, Davis PG, Schmidt B, Anderson PJ. Cognitive outcome at 24 months is more predictive than at 18months for IQ at 8- 9 years in extremely low birth weight children. *Early Hum Dev*. 2012;88:95-8.
- Schonhaut L, Schonstedt M, Álvarez J, Salinas P, Armijo I. Desarrollo psicomotor en niños de nivel socioeconómico medioalto. *Rev Chil Pediatr*. 2010;81:122-7.
- Guralnick MJ. Effectiveness of early intervention for vulnerable children: A developmental perspective. *Am J Ment Retard*. 1998;102:319-45.
- Anderson LM, Shinn C, Fullilove MT, et al. The effectiveness of early childhood development programs. *Am J Prev Med*. 2003;24:32-46.