



## APPLICABILITY OF THE BAYESIAN METHODOLOGY TO THE STUDY OF LOW INCIDENCE DISEASES: EXAMPLE OF CHILD ANXIETY

CONCEPCIÓN SAN LUIS<sup>1</sup>, GUILLERMO A. CAÑADAS<sup>5</sup>, JESÚS CANTERO<sup>2</sup>, LUIS M. LOZANO<sup>3</sup>,  
EMILIA I. DE LA FUENTE<sup>3</sup> y TATIANA LOZANO<sup>4</sup>

<sup>1</sup> Facultad de Psicología, Universidad Nacional de Educación a Distancia, Madrid

<sup>2</sup> Facultad de Medicina, Universidad de Granada, Granada

<sup>3</sup> Facultad de Psicología, Universidad de Granada, Granada

<sup>4</sup> Centro de Salud Goya 2, SERMAS, Madrid

<sup>5</sup> Escuela Universitaria de Ciencias de la Salud, Universidad de Granada, Granada

**Abstract:** This work is a cross-sectional quantitative study that provides information about anxiety symptoms from the CECAD questionnaire, applied to 500 students aged 8-12. Anxiety prevalence was estimated using three different procedures, both in the total sample and in a subsample of 10 children, randomly selected. Results showed a prevalence range of childhood anxiety for the three methods of 16.5%-23.5%, 16.7%-23.7%, and 18.3%-24%, when using the large sample. When the small sample size was used, the frequentist estimation method yielded impossible values, but results when using Bayesian methods were between 2.3% and 41.3% in the non-informative distribution, and between 17.9% and 27.4% in the informative distribution. The Bayesian method was found to be the procedure that provided better estimation, with improved results when informative distribution was used since it adds relevant information.

**Keywords:** Prevalence studies; child anxiety; low incidence; rare diseases; Bayesian empirical method.

*Aplicabilidad de la metodología bayesiana (método Bayes) en estudios de casos de baja incidencia:  
Ejemplo de la ansiedad infantil*

**Resumen:** Se presenta un estudio cuantitativo transversal que proporciona información sobre los síntomas de ansiedad evaluados mediante el cuestionario CECAD a 500 estudiantes de edades comprendidas entre los 8 y los 12 años. Se estimó la prevalencia de la ansiedad usando tres métodos estadísticos diferentes en el total de la muestra y en una submuestra de 10 estudiantes elegidos aleatoriamente del total. La prevalencia de ansiedad estimada por los tres métodos fue de 16.5%-23.5%, 16.7%-23.7% y 18.3%-24% en la muestra total. Cuando se emplea la submuestra, la estimación clásica (estadística "frecuentista") proporciona valores imposibles; empleando métodos bayesianos la estimación fue entre 2.3% y 41.3% con distribución inicial no informativa, y entre 17.9% y 27.4% con distribución inicial informativa. Los métodos Bayesianos proporcionan mejores estimaciones, siendo las estimaciones más precisas cuando se dispone de información inicial relevante.

**Palabras clave:** Estudios de prevalencia; ansiedad infantil; baja incidencia; enfermedades raras; método empírico Bayes.

Recibido: 7 julio 2010; aceptado 20 agosto 2010

*Corresponding author:* Concepción San Luis Costas, Universidad Nacional de Educación a Distancia, Departamento de Metodología y Ciencias de la Conducta, Facultad de Psicología, Juan del Rosal 10, 28040 Madrid. Correo-e: csanluis@psi.uned.es

*Acknowledgements:* This work has been founded with the projects SEJ2006-13009 of the Ministry of Education and Science (Spain) and the Project of Excellence P07HUM-02529 of the Junta de Andalucía (Spain).

During the last years, health sciences professionals have become more interested in improving the efficiency of quantitative and qualitative methods of research. Statistic procedures are the most used methods for data analysis and, amongst them, the interest in Bayesian methods is becoming more significant (Carlin & Louis, 2009; Travaglia, Westbrook & Braithwaite, 2009), because of the ability to interpret results

from very small data sets with greater precision than the frequentist methods.

Today, these methods are essential in the biomedical field in order to estimate the efficiency of possible treatments, survival parameters, long-term prognoses, and others. This is especially so in epidemiological sciences, that are ranging over a much larger field larger than before, focusing on the study of risk factors in the population with a view to developing population health policies (Hernández-Aguado, Lumbreras & Jarrin, 2006; Nettleton, Woods, Burrows & Kerr, 2009). Health professionals are particularly interested in developing methodologies for a better study of diseases, especially rare and low-incidence diseases, in general and in endemic cases.

A disease is considered rare when it is difficult to diagnose, and when, in the European Union, it has a prevalence of less than 5 cases per 10,000 inhabitants (Zurriaga-Llorens et al., 2006; European Commission, 2004). When a rare disease has to be studied, researchers usually have only a small number of cases. In these situations Bayesian methodology could offer more precise results (Congdon, 2003a,b; Guardia, De la Fuente & Lozano, 2008).

Recent studies show that emotional disorders in children are frequent (Carrasco-Ortiz & Del Barrio, 2007; Frías, Carrasco, Fernández, García & García, 2009), which appears to be directly related to their adult life mental health (Iglesias & Romero, 2009; Pilgrim, Rogers & Bentall, 2009). Anxiety is amongst the most frequent disorders in this population.

However, the results of many studies are not very consistent (Lozano, García-Cueto & Lozano, 2007) and sometimes contradictory due to, amongst other reasons: (a) studies may consider the terms “fear” and “anxiety” interchangeable and focus on the evolutive nature of specific fears, obviating many aspects of anxiety; (b) state and trait anxiety are not differentiated; (c) recognizing and quantifying internal symptoms may prove difficult (Sandín, Chorot, Valiente & Lostao, 2009; Sandin, Chorot, Valiente & Chorpita, 2010); (d) the psychometric properties of some evaluation instruments are deficient (Greenhill, Pine, March & Birmaher, 1998); (e) child anxiety may be incorrectly per-

ceived as a harmless and transient phenomenon (Benjamin, Costello & Warren, 1990). All these matters generate problems when trying to realistically estimate the prevalence of childhood anxious symptomatology. As a consequence, several authors (Bernstein & Borchardt, 1991; Costello et al., 1996), estimate a prevalence range of 12-20% for childhood anxiety disorders, 8% of cases possibly needing clinical treatment.

Another important reason for inconsistencies is that prevalence rates vary according to who is reporting on the child's anxiety (Berbel et al., 2010). A prevalence of 8% was found, based on parents' interviews. However, teacher reports put it at 21.2%, and direct child interviews resulted in 22.9% (Rivas, Vázquez & Pérez, 1995). The prevalence of anxiety versus other disorders in children requiring mental health services was 13.3%, with a male/female ratio of 0.69 (Aláez, Martínez-Arias & Rodríguez-Sutil, 2000).

Other authors reported a high comorbidity between anxiety and other mental disorders (Lozano et al., 2007; Royo, 2002), such as depression. In fact, it is difficult to differentiate between the two disorders (DSM IV-R names similar symptoms for anxiety and depression), with a usual coexistence in 70-80% of cases (Kovacs & Devlin, 1998). Therefore, the study of the two disorders requires a clear delimitation between them and estimation of the correspondent prevalences is subject to this delimitation, otherwise leading to apparently contradictory results.

There are different methodological procedures capable of estimating the prevalence of a disorder: the methods used in frequentist statistics and those proposed by the Bayesian statistics perspective. Frequentist statistics have been traditionally used in epidemiological research and the use of Bayesian methods is much more recent, but these have clear advantages over the first (Guardia, de la Fuente y Lozano, 2008).

The main advantages of the use of Bayesian methods are: (a) the previous information obtained in other studies can be used and implemented in the statistical analysis; (b) the Bayesian methods, unlike frequentist statistics, are particularly suited for small samples; (c) these analytical procedures do not require verification

of the parametric assumptions; (d) the accuracy of the parameter estimation is higher; (e) Bayesian credible intervals offer two boundaries for the parameter to be estimated with a certain probability, something that frequentist confidence intervals are erroneously said to offer; (f) two or more hypotheses can be tested; (g) the Bayesian hypotheses are solved calculating the probability of the hypothesis in issue. In relation to the advantages mentioned, it is worth noting that in case of non significant results, the probabilities of all the hypotheses in the study can be known (De la Fuente, Cañadas, Guàrdia & Lozano, 2009).

The aims of this study were to estimate the range of prevalence values for anxiety disorder in children aged 8-12, and to compare the results of the estimations using three method procedures, one frequentist method and two Bayesian methods, using two different sample sizes.

## METHOD

### *Participants*

A cross-sectional quantitative study was carried out. The sample consisted of 500 children in primary education, selected from the public and private schools of the Principado de Asturias (Spain), aged 8-12 (mean = 10.33 years, and Standard Deviation (SD) = 1.45). Participants were interviewed with the consent of the parents.

### *Measures*

All the children answered the CECAD (Clinical and Educational Questionnaire of Anxiety and Depression (Lozano et al., 2007), which includes values of anxiety, depression, uselessness, irritability, and thinking problems in Spanish children. However, in this work only the scores of the anxiety subscale were used. The questionnaire was applied by psychologists, specialized in group tests in children's classrooms. Questions were read to the younger children in order to ensure that they understood

them (as recommended in the questionnaire manual). After calculating the anxiety score of each child, anxious symptomatology was considered present in those with a score higher than the 95th percentile (Lozano et al., 2007).

### *Data analysis*

The statistical program SPSS 15.0 was used for frequentist data analysis. We used the maximum likelihood estimation of the parameter prevalence of anxiety ( $\hat{\theta}$ ). On the other hand, the WinBugs program was used (Spiegelhalter, Thomas, Best & Gilks, 1994; Spiegelhalter, Thomas, Best & Lunn, 2002; Guàrdia et al., 2008; Ntzoufras, 2009) for the Bayesian estimation of the prevalence of anxiety in order to compare results. We obtained the Bayes estimator of the prevalence without using information from previous research as well as the Bayes estimator which combines the previous information provided by Bragado (1995), with that contained in the data collected in this investigation. These options are called "non-informative" ( $\theta_{Non\ Inf}^*$ ) and "informative" ( $\theta_{Inf}^*$ ), respectively. Informative prior distribution of the prevalence of anxiety disorder was calculated using the method suggested by Congdon (2003a; 2003b) that parted from the 22.9% reported by Rivas, Vázquez, and Pérez (1995) for children aged 8-12. Beta distribution (1,1) was used as non-informative prior distribution.

A point and by interval estimation of the anxiety parameter was obtained using the three different techniques. Analyses were made for the total sample ( $N = 500$ ), and for a subsample of 10 children randomly selected, in order to study the divergence of estimations for a large or small sample size. For the Bayesian analyses (informative and non-informative), the quadratic loss function to obtain the point estimations was used.

### *Results*

Table 1 shows the results of the estimations using the three methods for the total sample ( $N = 500$ ), and the subsample ( $n = 10$ ).

Table 1. Point prevalence with confidence intervals of anxious symptoms in children aged 8-12, using three different estimation methods

<i>Estimation method</i>		<i>n = 10</i>	<i>N = 500</i>
Maximum likelihood estimation	Point estimation	0.100	0.200
	Inferior limit	-0.086	0.165
	Superior limit	0.286	0.235
Non-informative bayesian estimation	Point estimation	0.167	0.201
	Inferior limit	0.023	0.167
	Superior limit	0.413	0.237
Informative bayesian estimation	Point estimation	0.225	0.211
	Inferior limit	0.179	0.183
	Superior limit	0.274	0.240

Table 1 also shows that the lower limit of estimation using the frequentist statistical method gives negative values for the subsample of 10 children. This result is impossible, since the minimum possible value must be 0. This does not occur when Bayesian procedures (informative or non-informative) are applied. It is important that the interval amplitude (the difference between the lower and upper limit) is smaller when Bayesian estimation is used. The

smallest interval is obtained with the Bayesian estimation with initial information, giving this method the greater precision (see Figure 1).

These estimations are shown in Figure 1, which represents the point estimations of the prevalence of anxiety, using both the frequentist and Bayesian methods, in the two samples used (estimations in triangles, circles and squares, in black and white). The precision of the estimations is reflected by a segment around the cen-

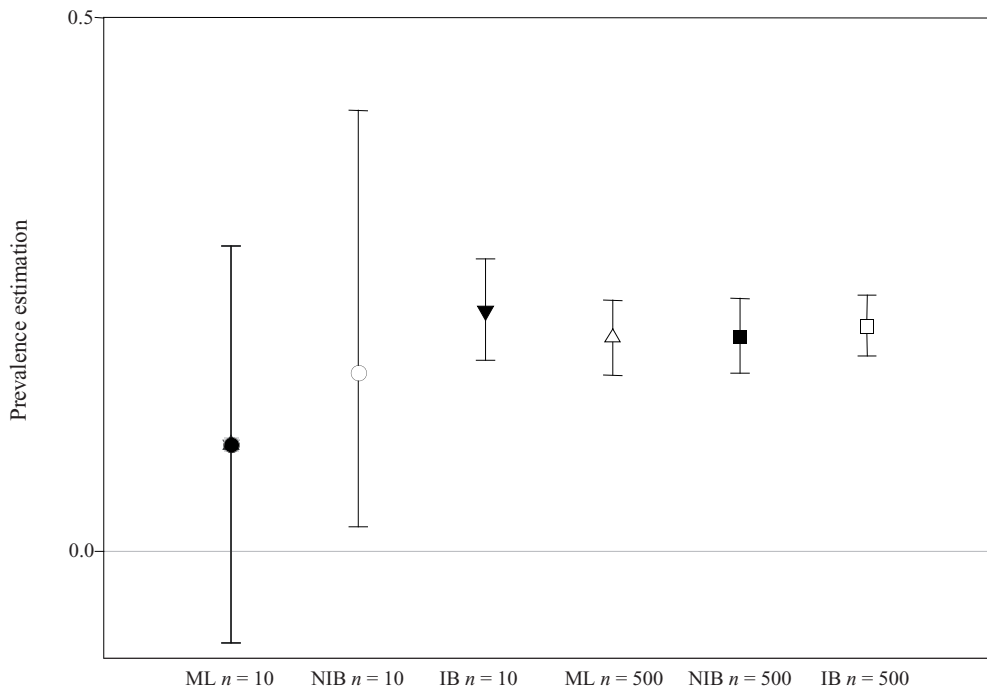


Figure 1. Point estimations of the prevalence of anxiety, using frequentist and Bayesian methods, in a subsample ( $n = 10$  children) and the total sample ( $N = 500$  children). ML= maximum likelihood; NIB= non-informative Bayesian method; IB= informative Bayesian method.

tral dot. As expected, Figure 1 shows that all methods lead to similar estimations and precisions of the disorder, when using the large sample ( $N = 500$ ). The remarkable finding came when a small sample ( $n = 10$ ) was used. Frequentist estimation not only provided impossible values in the interval estimation, but also produced very different estimations for the total sample ( $N = 500$ ) or the small sample ( $n = 10$ ).

The estimated prevalence using Bayesian procedures, however, produced very similar values for the large sample and the small sample. When informative Bayesian analysis was used, both estimations were found to be almost equal.

## DISCUSSION

The results of our study showed the prevalence of anxious symptoms in children aged 8-12 to be between 18.3% and 24%, with a point estimation of 21.1%. These high values are worrying and indicate that prevention programs for this disorder in childhood are needed. On the other hand, the results showed that Bayesian methods provided better estimations with small sample sizes and with greater precision.

It is worth noting that, as proposed by Bayarri and Cobo (2002), the possibility to integrate prior knowledge with empirical information obtained from other data sets improved the results considerably, offering a 18.6% more accurate estimation than the classical confidence interval.

Health sciences research often has to work with small samples. Thus an important result of the study is that in the randomly selected small sample, of only 2% of the total sample, the Bayesian estimations were found to be more precise than the frequentist method, especially with the prior information method.

Therefore, Bayesian methodology appears a more than viable option in the study of rare and low incidence diseases (Hernández-Aguado et al., 2006; Zurriaga Llorens et al., 2006). Prior information should be used to improve the efficiency of the method and, therewith, the results (Bayarri & Cobo, 2002; Álamo, Vázquez & Rodríguez, 2002).

Frequentist statistical procedures may produce erroneous results when only very small samples are available, because the asymptotic theory underlying these techniques requires large sample sizes.

Knowing the magnitude of anxiety prevalence in children is important for establishing preventive programs. The aim of these programs should be twofold: in the first place, to reduce anxiety, and, in the second place, to teach emotional education to avoid the appearance of high anxiety levels. Individualized medical and psychological assessment to differentiate between children with symptoms of anxiety and those with clinical anxiety disorders is recommended.

## REFERENCES

- Alález, M., Martínez-Arias, M. R., y Rodríguez-Sutil, C. (2000). Prevalencia de trastornos psicológicos en niños y adolescentes, su relación con la edad y género. *Psicothema*, *12*, 525-532.
- Álamo, F., Vázquez, F. J., y Rodríguez, J. C. (2002). Herramientas para la investigación Biomédica: La perspectiva Bayesiana (I y II). *Medicina Clínica*, *119*, 265-272.
- Bayarri, M. J., y Cobo, E. (2002). Una oportunidad para Bayes. *Medicina Clínica*, *119*, 252-253.
- Benjamin, R. S., Costello, A. J., & Warren, M. (1990). Anxiety disorders in pediatric sample. *Journal of Anxiety Disorders*, *4*, 293-316.
- Berbel, E., Sepúlveda, A.R., Graell, M., Andrés, P., Carrolles, J. A., y Morandé, G. (2010). Valoración del estado de salud y psicopatología de los familiares en el trastorno del comportamiento alimentario: Diferencias entre cuidadores principales y secundarios. *Revista de Psicopatología y Psicología Clínica*, *15*, 179-192.
- Bernstein, G. A., & Borchardt, C. M. (1991). Anxiety disorders of childhood and adolescence: A critical review. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*, 519-532.
- Bragado, C. (1995). Prevalencia de los trastornos psicológicos en niños y adolescentes: resultados preliminares. *Clínica y Salud*, *6*, 67-82.
- Carlin, B. P., & Louis, T.A. (2009). *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC.
- Carrasco-Ortiz, M. A., & Del Barrio, V. (2007). Temperamental and Personality variables in child and adolescent depressive symptomatology. *Psicothema*, *19*, 43-48.
- Congdon, P. (2003a). *Applied Bayesian modelling*. New York: Wiley.
- Congdon, P. (2003b). *Bayesian statistical modelling*. New York: Wiley.

- Costello, E. J., Angold, A., Burns, B. J., Stangl, D. K., Tweed, D. L., Erkanli, A., & Worthman, C. M. (1996). The great smoky mountains study of youth: goals, design, methods and the prevalence of DSM-III-R disorders. *Archives of General Psychiatry*, *53*, 1129-1136.
- De la Fuente, E.I., Cañadas, G.R., Guàrdia, J. y Lozano, L.M. (2009). Hypothesis probability or statistical significance? Inference for the mean in a normal distribution. *Methodology*, *5*, 35-39.
- European Commission. (2004). *Health and consumer protection directorate-general. Useful information on rare diseases from an EU perspective*. Luxembourg: European Commission.
- Frías, A., Carrasco, A., Fernández, M. J., García, A., y García, B. (2009). Comorbilidad entre trastornos de conducta y depresión mayor en la adolescencia: Prevalencia, teorías explicativas y estatus nosológico. *Revista de Psicopatología y Psicología Clínica*, *14*, 1-16
- Greenhill, L. L., Pine, D., March, J., & Birmaher, M. (1998). Assessment issues in treatment research of pediatric anxiety disorders: what is working, what is not working, what is missing and what needs improvement. *Psychopharmacol Bulletin*, *34*, 155-164.
- Guàrdia, J., De la Fuente, E. I., y Lozano, L. M. (2008). Bayesian Inference for Binomial Populations. Bayesian Estimation for Child Depression Prevalence. *Advances and Applications in Statistics*, *9*, 13-35.
- Hernández-Aguado, I., Lumbreras, B., y Jarrin, I. (2006). La epidemiología en la salud pública del futuro. *Revista Española de Salud Pública*, *80*, 469-474.
- Iglesias, B., y Romero, E. (2009). Estilos parentales percibidos, psicopatología y personalidad en la adolescencia. *Revista de Psicopatología y Psicología Clínica*, *14*, 63-77.
- Kovacs, M., & Devlin, B. (1998). Internalizing disorders in childhood. *The Journal of Child Psychology and Psychiatry*, *39*, 47-63.
- Lozano, L., García-Cueto, E., y Lozano, L. M. (2007). CECAD. *Cuestionario educativo clínico: ansiedad y depresión*. Madrid: TEA-Ediciones.
- Nettleton, S., Woods, B., Burrows, R., & Kerr, A. (2009). Food allergy and food intolerance: Towards a sociological agenda. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, *13*, 647-664.
- Ntzoufras, I. (2009). *Bayesian Modelling using WinBugs*. New York: Wiley.
- Pilgrim, D., Rogers, A., & Bentall, R. (2009). The centrality of personal relationships in the creation and amelioration of mental health problems: The current interdisciplinary case. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, *13*, 235-254.
- Rivas, M.V., Vázquez, J., y Pérez, M. (1995). Alteraciones psicopatológicas en la infancia: el niño en E.G.B. *Psicothema*, *7*, 513-526.
- Royo, J. (2002). Trastornos depresivos en la infancia y adolescencia. *Anales del Sistema Sanitario de Navarra*, *25* (Supl.3), 97-103.
- Sandín, B., Chorot, P., Valiente, R. M., & Chorpita, B. F. (2010). Development of a 30-item version of the Revised Child Anxiety and Depression Scale. *Revista de Psicopatología Clínica*, *15*, 165-178.
- Sandín, B., Valiente, R. M., y Chorot, P. (2009). RCADS: Evaluación de los síntomas de los trastornos de ansiedad y depresión en niños y adolescentes. *Revista de Psicopatología y Psicología Clínica*, *14*, 193-206.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. (1994). *Bugs Examples*. MRC Biostatistics Unit, Cambridge 2002. Retrieved April 28, 2009, from <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/vol1.pdf>.
- Spiegelhalter, D. J., Thomas, A., Best, N. G., & Lunn, D. (2002). *WinBugs v. 1.4 User Manual*. MRC Biostatistics Unit, Cambridge. Retrieved April 28, 2009, from <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/manual14.pdf>.
- Travaglia, J.F., Westbrook, M.T. & Braithwaite, J. (2009). Implementation of a patient safety incident management system as viewed by doctors, nurses and allied health professionals. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine*, *13*, 277-296.
- Zurriaga, Ó., Martínez, C., Arizo, V., Sánchez, M. J., Ramos, J. M., García, M. J., Ferrari, M. J., et al. (2006). Los registros de enfermedades en la investigación epidemiológica de las enfermedades raras en España. *Revista Española de Salud Pública*, *80*, 249-257.