Evaluation size and number of yellow potato tubers under different planting densities using zero-inflated models

Evaluación del tamaño y número de tubérculos de papa amarilla bajo diferentes densidades de siembra usando modelos cero inflados

Avaliação do tamanho e número de tubérculos de batata amarela sob diferentes densidades de plantio usando modelos inflados a zero

Nelson Bernal-Margfoy
Enrique Darghan
Luis Ernesto Rodríguez
German Gonzalez*

Facultad de Ciencias Agrarias, Universidad Nacional de Colombia, sede Bogotá, www.unal.edu.co, Bogotá, Colombia.

Rev. Fac. Agron. (LUZ). 2022, 39(4): e223949
ISSN 2477-9407
DOI: https://doi.org/10.47280/RevFacAgron(LUZ).v39.n4.04

Received: 09-08-2022
Accepted: 27-09-2022
Published: 25-10-2022

Crop Production
Associate editor: Dr. Jorge Vilchez-Perozo
University of Zulia, Faculty of Agronomy
Bolivarian Republic of Venezuela

Keywords:
Zero-inflated model
Size tubers
Negative binomial regression

Abstract

A field study was carried out on the cultivation of the yellow diploid potato (*Solanum tuberosum* Phureja Group) to evaluate the influence of the planting density associated with distances between plants of 30, 40, and 50 cm and distances between the rows and paths of 100 cm on the tuber count with sizes less than 2 cm, 2-4 cm, 4-6 cm, and more than 6 cm. At the time of the harvest of the tubers, they were classified by size and respective count was made. The modelling of the counts was done by means of the usual negative binomial regression and by the inflated zeros option. The zero-inflated negative binomial regression models showed a significant effect of the sowing density on the tuber count in the sizes that were superior to 4 cm while the negative binomial model showed a significant effect it in the case of the sizes lower than 4 cm. The results on size and density are attributes of interest in both the agronomic management of this crop and in the industrial management of the tubers, so the relationship that we found can be adopted in both areas to generate the desired attributes of the crop for improving the production and guiding the process of industrialization.

*Corresponding author: gdgonzalezs@unal.edu.co

Resumen

Se realizó un estudio de campo en el cultivo de papa diploide amarilla (*Solanum tuberosum* Grupo Phureja) para evaluar la influencia de la densidad de siembra asociada a distancias entre plantas de 30, 40 y 50 cm y distancias entre hileras y caminos de 100 cm en el número de tubérculos con tamaños menores de 2 cm, 2-4 cm, 4-6 cm y mayores de 6 cm. Al momento de la cosecha de los tubérculos, se clasificaron por tamaño y se realizó el conteo respectivo. El modelado de los conteos se realizó mediante la habitual regresión binomial negativa y mediante la opción de ceros inflados. Los modelos de regresión binomial negativa con ceros inflados mostraron un efecto significativo de la densidad de siembra sobre el conteo de tubérculos en las tallas superiores a 4 cm mientras que el modelo binomial negativo mostró un efecto significativo en el caso de las a 4 cm. Los resultados sobre tamaño y densidad son atributos de interés tanto en el manejo agronómico de este cultivo como en el manejo industrial de los tubérculos, por lo que la relación en contrada puede ser adoptada en ambas áreas para generar los atributos deseados del cultivo para mejorar la producción y orientar el proceso de industrialización.

Palabras clave: modelo inflado con ceros, tamaño de tubérculos, regresión binomial negativa.

Resumo

Foi realizado um estudo de campo no cultivo da batata diplóide amarela (*Solanum tuberosum* Grupo Phureja) para avaliar a influência da densidade de plantio associada a distâncias entre plantas de 30, 40 e 50 cm e distâncias entre linhas e caminhos de 100 cm na contagem de tubérculos com tamanhos inferiores a 2 cm, 2-4 cm, 4-6 cm e superiores a 6 cm. No momento da colheita dos tubérculos, os mesmos foram classificados por tamanho e foifeita a respectiva contagem. A modelagem das contagens foi feita por meio da usual regressão binomial negativa e pela opção de zeros inflados. Os modelos de regressão binomial negativa inflados a zero mostraram efeito significativo da densidade de semeadura sobre a contagem de tubérculos nos tamanhos superiores a 4 cm enquanto o modelo binomial negativo mostrou efeito significativo no caso dos tamanhos inferiores a 4 cm. Os resultados de tamanho e densidade são atributos de interesse tanto no manejo agronômico desta cultura quanto no manejo industrial dos tubérculos, de modo que a relação que encontramos pode ser adotada em ambas as áreas para gerar os atributos desejados da cultura para melhorar o produção e orientando o processo de industrialização.

Palavras-chave: modelo inflado de zero, tamanho de tubérculos, regressão binomial negativa.

Introduction

The Colombian yellow diploid potato is an important genetic resource, consisting primarily of diploid genotypes with short-day adaptation. These potatoes are of economic importance in Andean countries, especially since they are cultivated by small farmers. Colombia is the largest producer, exporter and consumer of the yellow diploid potatoes (Barragán, 2019). The production of yellow potatoes is classified in different categories according to the diameter of the tuber, for example, diameters less than 2 cm, 2-4 cm and diameters greater than 4 cm (Tabares *et al*. 2009).

There are various investigations relating the size and quality of yellow potato tubers to planting density. In the Bogotá savannah in Colombia, under these agro-climatic conditions, high planting densities (0.7 m between rows and 0.2 m between plants) favour the increase of small tubers (Arias *et al*. 1996). The same authors relate the stem density with the number of tubers produced in different planting densities and conclude that the stem density increases with an increase in the planting density. A higher planting density is related to a higher yield per unit area, but lower yield per plant, showing consistency between the planting densities tested for all trials in a study in subtropical China (He *et al*. 1998).

According to Rodríguez *et al*. (2003), it is important to determine the effects of the density components (represented mainly by the size of the tuber-seed, the spacing, the number of stems per plant and, in general, the spatial arrangement) on the morphology, physiology and the agronomic performance of the cultivar in different environmental conditions.

Appropriate modelling can yield contrasting results with those analyses where the data has been intervened to find the assumptions for the application of a specific technique, so that the choice of an appropriate methodology could yield different results from those where the initial distribution has been modified; and, therefore, the absence of statistical significance due to planting density can not only be attributed to the factors involved but to the model used (Wasserstein and Lazar, 2016)

In some agronomic research it is possible to find count data (enumerated units) with additional zeros from those that would be expected in those situations. This is common in the modelling of counts with excess zeros or inflated by zeros (Hilbe 2014).

In this research, the modelling of the number of tubers per size (four categories) is proposed using regression models by number, specifically the zero-inflated negative binomial model using three planting densities as the only factor in two repetitions per density and with sufficient experimental units to avoid problems in estimating the over-dispersion parameters due to the sample size.

Materials and methods

This research was carried out at the Marengo Agricultural Centre of the Universidad Nacional de Colombia, Mosquera municipality, Cundinamarca department (74°12'58.51 W; 4°40'52.92 N), at an altitude of 2516 meters above sea level. The mean temperature is 14 °C in a range from 12 °C to 18 °C and mean rainfall of 500 mm to 1000 mm. The soils are moderately deep and well drained, and the water table is less than 0.5 m below the surface with a 15% moisture content. According to the characteristics of precipitation, temperature and evapotranspiration, the area is classified as low montane dry forest.

Yellow diploid potato seeds (*Solanum tuberosum* Phureja Group) cv. Criolla Colombia are sown one per planting site, four centimetres in diameter, without rotting or skin defects. This cultivar is yellow, has a medium plant size, with slightly light green foliage and abundant flowering. Tuber production shows a size distribution of diameters between 1 and 10 cm. The tubers are harvested at 120 days and counted according to their diameter in the categories 2, 2-4, 4-6 and > 6 cm. To study the effect of planting density, the plants were planted at 30, 40 and 50 cm between plants, all with 100 cm separation between rows (figure 1).

Figure 1. Harvested tubers of different sizes.

Sowing was carried out in rows precisely aligned according to the planting density, using three successive rows according to the geometry of the lot for each density with two repetitions per density for a total of 18 rows and 2841 plants. The experimental unit was the plant (tubers), sown on different plots within the same lot (blocks), recording the data for each plant. Under these conditions, the design turned out to be a simple factorial in a generalized randomized block (complete) design, taking the distances between plants as the factor levels associated with the plant density.

The statistical analysis initially involved the descriptive component, generating cross tables and plots for the number of tubers by density and size. For the inferential component various models of the generalized linear type were adjusted, finally presenting those with the best descriptors of the adjustment and fulfilment of assumptions. From the exploratory analysis, it was already known that the negative binomial distribution seemed to generate the best fit, so the following regression models used the negative binomial distribution for each size, using the planting density as a factor. The analysis was carried out in the R software and the package used was Hilbe JM (2014). COUNT: Functions, Data and Code for Count Data. R package version 1.3.4.

Results and discussion

The size of the tubers seems to be regulated by various mechanisms and Struik *et al*. (1990) consider the planting density, the number of stems per plant, and the number of tubers per stem as important variables in the manipulation and modelling of crop yield. Cotes *et al*. (2000) recommend for this cultivar a size of less than 2 cm but without the effect of the planting density. However, they recommend a distance between plants of 30 cm with 130 cm between rows. The study of the size of the tubers plays an important role in achieving a presentation and conservation of the tubers that is acceptable to consumers and industry. In the case of potatoes for freezing or precooking, low sizes are desirable; and in the case of fried tubers or strips, sizes greater than 4 cm are desirable. In another study, the decrease in planting density increased the number of tubers per seedling for all sizes. In fact, a reduction from approximately 146 to 25 plants.m^2 doubled the number of usable tubers (Van der Veeken and Lommen 2009).

Considering these ideas, the results shown below are associated with modelling the number of yellow potato tubers for three different planting densities and for four sizes of tubers. Table 1 represents the number of tubers by size and the total number for each planting density, the highest number is in the distances between plants of 30 cm and 40 cm. However, the percentages of tubers conditioned by density are similar in the three densities, which at first glance suggests that there are no important differences in the number of tubers obtained by density.

Table 1. Distribution of the number of tubers by density and size class marks.

Density	Sizes (cm)				Sub-total
	≤ 2	(2 - 4]	(4 - 6]	> 6	
d1:(30*100cm)	8258	8337	4152	159	20906
d2:(40*100cm)	8088	8538	4254	215	21095
d3:(50*100cm)	7291	7064	4406	220	18981
Sub-total	23637	23939	12812	594	**Total (60982)**

In figure 2 the behaviour of the number of tubers is illustrated discriminated by density, to show the excesses of zeros present in the two largest sizes for all the planting densities considered. The sizes were labelledS1, S3, S5, and S9 and only indicate the midpoint of the sizes used.

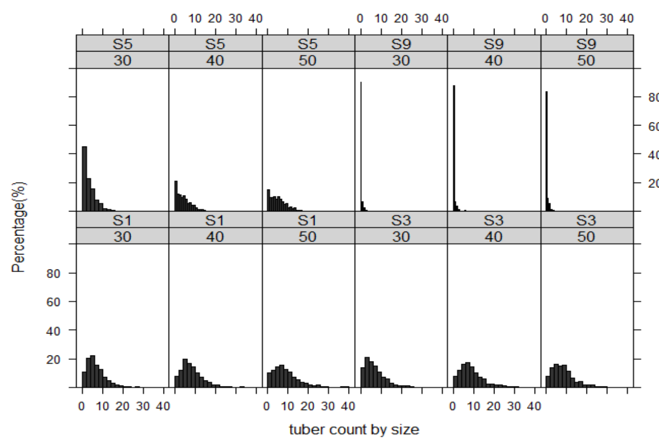

Figure 2. Comparison of observed and predicted values by size and density.

Several studies include the two most used distributions for this type of data, namely the Poisson and the negative binomial or its zero-inflated option (Henne, 2012). For this reason, below, the results obtained by some test models and the models finally proposed are illustrated to relate the number of tubers with relation to the planting density using the negative binomial regression model, which, through the graphs previously shown seemed to be adequate in the case of modelling the sizes for each density. Modelling was done for sizes, since the presence of excessive zeros in the largest sizes was evident, therefore, from now on, the results of the modelling with negative binomial regression and zero inflated negative binomial are presented.

In the negative binomial model, the variance is not restricted to being equal to the mean and facilitating the modelling like the data we are evaluating, where the variance is a little more than the triple the average. The negative binomial allowed the manipulation of excesses of zeros as they are found in the larger diameters or sizes, so this distribution is quite justified for modelling the numbers of tubers by density. The data collected showed an excess of zeros especially in the two upper sizes, so it was necessary to contrast this by means of a test, the negative binomial models with their counterpart that accepts excess zeros. Next, the results obtained using the R software (pscl package) for the construction of the outputs associated with the estimation and adjustment process of the zero-inflated negative binomial models are described using density as a factor and modelling the response by size.

Tables 2 and 3 show the results of the analysis, and these two cases are shown initially since they are the ones that did not have an evident excess of zeros in the number of tubers associated with the zero value, in this case the usual negative binomial model was used. In both cases the significant effect of planting density on size is evident, as found in various studies that evaluate planting density as a factor that presumably affects yield. However, we cannot assert anything with respect to the similarity with other research, since this is not usually the form of modelling found in the literature for counts.

Table 2. Negative binomial model (log link) by density at size ≤ (2 cm).

Coefficient	Estimation	Standard Error	Z-value	Probability
Intercept	1.985	0.019	103.87	$2.00e^{-16}$
Density (40)	1.183	0.028	6.51	$7.50e^{-11}$
Density (50)	0.253	0.029	8.64	$2.00e^{-16}$

Table 3. Negative binomial model (log link) by density in size (2-4) cm.

Coefficient	Estimation	Standard Error	Z-value	Probability
Intercept	1.994	0.020	101.36	$2.00e^{-16}$
Density (40)	0.227	0.029	7.88	$3.20e^{-15}$
Density (50)	0.212	0.030	7.00	$2.70e^{-12}$

The results of tables 4 and 5 show again the significant effect of the terms associated with the planting density on the number of tubers. In both tables the negative binomial was used with the logarithmic link function. The two parts of an inflated model represent a binary model, generally a logit model for which of the two processes the zero result is associated and a count model (in this case, a negative binomial model). In both cases the statistical relationship between the density of seeding and the count of potato tubers was evident.

Table 4. Zero Inflated negative binomial(log link) for density for the sizes 4-6 cm.

Coefficients	Estimation	Standard Error	Z Value	Probability
Intercept	1.445	0.026	56.43	$2.00e^{-16}$
Density (40)	0.163	0.036	4.51	$6.50e^{-6}$
Density (50)	0.349	0.037	9.53	$2.00e^{-16}$
Theta	1.180	0.062	19.14	$2.00e^{-16}$

Table 5. Zero Inflated negative binomial (log link) in the sizes of more than 6 cm.

Coefficients	Estimation	Standard Error	Z Value	Probability
Intercept	-0.870	0.266	-3.26	0.0011
Density (40)	0.792	0.219	3.61	0.0003
Density (50)	0.641	0.219	2.93	0.0034
Log(theta)	0.217	0.515	0.42	0.6740

With all these results, it only remains to graphically show the results of the adjustment of the first two sizes in the negative binomial models and the two upper sizes in the negative binomial models inflated by zeroes. Figure 3 corroborates the adjustment obtained in the Poisson, negative binomial (Neg Bin) and negative binomial adjustment modalities inflated by zeros (Zinb). In all cases observed for each size and density (Observed 1: sizes ≤ 2 cm (A); Observed 3: sizes (2-4) cm (B); Observed 5: sizes (4-6) cm (C) and Observed 9: size > 6 cm (D) the models in the negative binomial distribution were superior; however, the inflated option by zeros was the most convenient in the two larger sizes.

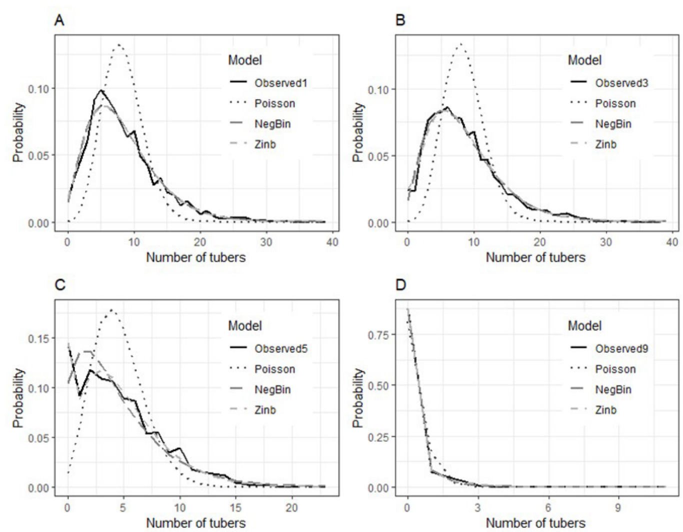

Figure 3. Comparison of observed and predicted values by size and density.

This scientific publication in digital format is a continuation of the Printed Review: Legal Deposit pp 196802ZU42, ISSN 0378-7818.

Once the modelling process has concluded and the adequacy of the modelling is clearly appreciated respecting the nature of the variables, it can be statistically asserted that the planting density factor explains the numbers for all the sizes evaluated, corroborated by different authors like Escobar and Zaag (1988), where an increase in the sowing density from 40,000 to 100,000 plants per hectare increased the yield by 50%; but smaller tubers were generated, interpreted as an effect of the density on the size of the tubers.

The results of table 6 suggest the need to use low planting densities (fewer plants per hectare) if the market requires larger tubers. However, this is accompanied by low yields that will most likely be offset by higher costs when selecting tubers by size. Note that average tubers per plant is reduced by almost 130% in the two extreme densities. Additionally, in the first density for every 27 tubers of the two largest sizes approximately 104 of the smaller sizes are generated (almost 4:1), whereas in the lowest density the ratio is approximately 3:1 $((33+32).(20+1))^{-1}$.

Table 6. Distribution of fresh weight ($t.ha^{-1}$), mean of tubers per plant and ratios of generated tuber by size in each density.

Density	Fresh weight ($t.ha^{-1}$)	$Tubers.plant^{-1}$	Ratio
d1:(30cm*100cm)	10.77	18.4	52:52:26:1
d2:(40cm*100cm)	8.20	11.4	38:40:20:1
d3:(50cm*100cm)	5.37	8.1	33:32:20:1

Tuberization in potatoes involves the differentiation of stolon in young tubers (initiation) and the collection of young tubers (Dutt *et al*. 2017). Competition for resources at high densities can affect tuberization by reducing the number of starting tubers (Mackerron *et al*. 1988). In addition, these resourced-related stresses (for example water) can reduce tuber filling with assimilated tubers in the plant's growth phase (Lahlou *et al*. 2003). In both cases, the result in a reduction in tuber yield.

Marketable tuber yield depends on the average tuber size, that is, both the total tuber weight and the total number of tubers. Therefore, cultivars that produce fewer tubers in drought-prone areas are recommended. If you have a smaller number of tubers, it is more likely that they are larger when the photo-assimilated are limited during drought, thus increasing their average size (Aliche *et al*. 2019)

The negative binomial distribution or zero-inflated negative binomial model can provide information on the marketable proportion of tuber yields. However, not much research has been conducted towards understanding the underlying reason for the model parameters that describe total and marketable tuber size distribution, although it seems to be associated with the number and size of tubers under quantitative inheritance (Celis-Gamboa, 2002). Despite this, the relationship between the density of seeding and the count of tubers by size was evident and can be used to direct the production in favour of generating the sizes required by the market.

Conclusions

In sizes less than 4 cm adjusting negative binomial models without excess zeros found that the terms associated with the planting density were more appropriate to show the statistical relationship between the density of the seedlings and the number of tubers. Similarly, in sizes greater than 4 cm adjusting negative binomial models with excess zeros showed the terms associated with the sowing density were the ones with the best statistical adjustment. So, it can be statistically asserted that sowing density influences the number of tubers in larger sizes.

Larger tuber sizes were associated with lower planting density, but this was associated with lower yields, suggesting that there is a yield penalty in the interest of improving tuber sizes.

Literature cited

Aliche, E. B., Oortwijn, M., Theeuwen, T. P. J. M., Bachem, C. W. B., van Eck, H. J., Visser, R. G. F. & van der Linden, C. G. (2019). Genetic mapping of tuber size distribution and marketable tuber yield under drought stress in potatoes. *Euphytica*, *215*(11). https://doi.org/10.1007/s10681-019-2508-0

Arias, V., Bustos, P. & Ñustez, C. E. (1996). Evaluation of the yield in papa criolla (*Solanum phureja* Juz. et BUk.), «yema de Huevo» variety, under different densities at Sabana de Bogotá. *Revista Agronomía Colombiana, 13*(2), 152-161. https://revistas.unal.edu.co/index.php/agrocol/article/view/25442

Barragán, J. N. (2019). Desempeño y perspectivas económicas del subsector papa 2018-2019. Revista Papa (47), 44-48.

Celis-Gamboa, B. C. (s. f.). *The life cycle of the potato (Solatium tuberosum* L.*): From crop physiology to genetics* [Ph.D. Thesis]. Wageningen University. https://www.proquest.com/openview/96d595f31eaa4257cbf73c2041988475/1?pq-origsite=gscholar&cbl=2026366&diss=y

Cotes, J. M., Ñústez, C. E. & Pachón, J. I. (2000). Evaluación de la densidad de siembra y el tamaño del tubérculo semilla en la producción de semilla básica de papa criolla, variedad yema de huevo (*Solanum phureja* Juz et Buk.). *Revista Agronomía Colombiana*, *17*(1-3), 57-60. https://revistas.unal.edu.co/index.php/agrocol/article/view/21546/22550

Dutt, S., Manjul, A. S., Raigond, P., Singh, B., Siddappa, S., Bhardwaj, V., Kawar, P. G., Patil, V. U. & Kardile, H. B. (2017). Key players associated with tuberization in potato: potential candidates for genetic engineering. *Critical Reviews in Biotechnology*, *37*(7), 942-957. https://doi.org/10.1080/07388551.2016.1274876

Escobar, V. and Vander Zaag, P. (1988). Field performance of potato (*Solanum* spp.) cuttings in the warm tropics: Influence of planting system, hilling, density and pruning. *American Potato Journal*, *65*(1), 1-10. https://doi.org/10.1007/bf02855308

He, W., Struik, P. C., He, Q. & Zhang, X. (1998). Planting Time and Seed Density Effects on Potato in Subtropical China. *Journal of Agronomy and Crop Science*, *180*(3), 159-171. https://doi.org/10.1111/j.1439-037x.1998.tb00386.x

Henne, D. C., Workneh, F. & Rush, C. M. (2012). Spatial Patterns and Spread of Potato Zebra Chip Disease in the Texas Panhandle. *Plant Disease*, *96*(7), 948-956. https://doi.org/10.1094/pdis-09-11-0805-re.

Hilbe, J. M. (2014). Modeling Count Data. *Cambridge University Press*. https://doi.org/10.1017/cbo9781139236065

Lahlou, O., Ouattar, S. & Ledent, J. F. (2003, abril). The effect of drought and cultivar on growth parameters, yield and yield components of potato. *Agronomie*, *23*(3), 257-268. https://doi.org/10.1051/agro:2002089

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rodríguez Pérez, L., Corchuelo Rozo, G. & Ñústez, C. E. (2003). The influence of plant spacing on potato (*Solanum tuberosum* L. cv. Parda pastusa) morphology and growth in two contrasting environments. *Revista Agronomía Colombiana*, *21*(3), 210-21. http://www.redalyc.org/articulo.oa?id=180317974010

Struik, P. C., Haverkort, A. J., Vreugdenhil, D., Bus, C. B. & Dankert, R. (1990). Manipulation of tuber-size distribution of a potato crop. *Potato Research*, *33*(4), 417-432. https://doi.org/10.1007/bf02358019

Tabares Patiño, E. D., Jaramillo Villegas, S., González Santamaría, L. H. & Cotes Torres, J. M. (2009). Effect of fertilization in potato (*Solanum tuberosum* L.) DIACOL Capiro variety under Andisol from EAST Antioquia, COLOMBIA. *Revista Facultad Nacional de Agronomía - Medellín*, *62*(2), 5099-5110. http://www.scielo.org.co/pdf/rfnam/v62n2/a09v62n2.pdf

van der Veeken, A. J. H. and Lommen, W. J. M. (2009). How Planting Density Affects Number and Yield of Potato Minitubers in a Commercial Glasshouse Production System. *Potato Research*, *52*(2), 105-119. https://doi.org/10.1007/s11540-008-9124-z

Wasserstein, R. L. & Lazar, N. A. (2016). The ASA Statement on *p*-Values: Context, Process, and Purpose. *The American Statistician*, *70*(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108