

Geospatial tools and Chagas control: a critical comment on the paper of Weinberg *et al.*, Evaluation and planning of Chagas control activities using geospatial tools. *Geospat Health* 2019;14:229-238

David E. Gorla

Instituto de Diversidad y Ecología Animal, CONICET – Universidad Nacional de Córdoba, Argentina

The contribution of the geospatial tools to the vector control activities of Chagas disease is very clear for the academic and technology-prone communities, but not so much for the state-dependent health agents (national or provincial) that are ultimately responsible for disease control in Latin America. In spite of many published scientific papers during the last 20 years on the use of geospatial tools for the surveillance and control of Chagas disease, that I started contributing to more than 20 years ago (Gorla *et al.*, 1997), health systems did not widely adopt these tools in their daily routines. The lack of technically skilled personnel within the health agencies may have been one of the causes of the adoption failure. In some other cases, one can identify products of the academic/technical communities that do not give enough attention to study designs that in the end discourages the end-users to adopt the tools.

The article by Weinberg *et al.* (2019) on the use of geospatial tools in the control of Chagas disease is just one example of a good intentions that probably will not help health agencies adopt the geospatial tools needed to improve the performance of the field programs because of what I perceive as a weak study design. With this comment, I invite the scientific and technical community to make a critical appraisal of what we do through a particular consideration of the mentioned article, although I suspect it applies to our study area in general.

One of the main problems in the mentioned article is trying to squeeze data to provide information without a clear study design. The authors inform that although “there was no experimental design”, they “tried to capture the information contained in control

variations estimated through spatial indices”, including as the main conclusion that “tools can be developed and introduced to extract information from existing data”. Observational studies without a clear study design, no matter how sophisticated the quantitative analytical methodology used, will run the risk of making misleading interpretations. The usual argument supporting this flawed approach is that it answers the question “how much explanation of the variability in the response variable can be obtained considering only the spatial effects?”. The problem with this argument is that in complex systems such as Chagas, where multi causality is the norm, considering only part of the system could lead to misleading (or false) conclusions. The risk of a partial analysis of a complex system is increased when there is weak study design, and the whole study is under question when the primary data quality is low.

In the article by Weinberg *et al.* (2019), the response variable is house infestation by *Triatoma infestans* (fraction of evaluated houses harbouring vectors), estimated with the man-hour active search technique. Although the technique is the standard protocol recommended by the Pan American Health Organization (PAHO), it is widely known that it has low sensitivity (especially when vector population abundance is low), giving many false negatives (Abad-Franch *et al.*, 2014). A clear warning on the problem of imperfect detection was published a decade ago by Abad-Franch and Ferraz (2010). So, analysing this dataset as it were the truth gives a false impression of certainty and contest all the construction built over it.

Evaluating the effect of the Integral Chagas Program referred to in the article on the response variable, based mainly on spatial effects without consideration of other factors we know do have effect, namely i) rural-urban migration; ii) effect of the history of land-use change (very strong in the Añatuya area, as seen by the agricultural patchwork of the Sentinel image included in the article, that did not exist 20 years ago); iii) effects of house quality/house improvement (house improvement sometimes works as generally expected according to the accepted knowledge, but sometimes it does not, as shown by Cavallo *et al.*, 2018), will seriously affect the interpretation of the analysis.

Inaccuracies in Weinberg’s article do not help to transmit a clear message on the use of geospatial tools. “Chagas control” as used by the authors refers to “vector control activities”, as no data on the disease were considered. The statement “reports evaluating control programmes that consider the spatiotemporal history of infestation and spraying of a site are generally lacking in the literature” is not quite true. To name just a few articles that started appearing 15 years ago on spatiotemporal analysis of *T. infestans* infestations: Cecere *et al.* (2004), Zu Dohna *et al.* (2007, 2009), Levy *et al.* (2010), Provecho *et al.* (2017), Cavallo *et al.* (2018).

Correspondence: David E. Gorla, Instituto de Diversidad y Ecología Animal, CONICET – Universidad Nacional de Córdoba, Argentina
E-mail: david.gorla08@gmail.com.

Conflict of interest: The Authors declare no potential conflict of interests.

Received for publication: 26 May 2020.
Accepted for publication: 26 May 2020.

©Copyright: the Author(s), 2020
Licensee PAGEPress, Italy
Geospatial Health 2020; 15:901
doi:10.4081/gh.2020.901

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License (CC BY-NC 4.0) which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.



Much more effort should be dedicated to the design of the study. In this case, distance between houses (either infested or sprayed with insecticides) is considered as the factor driving the variation of the response variable, as authors say “This could indicate that in order to be free from infestation, a spraying at least 300 m away from each infested house would have to be guaranteed”. Distance to/from a house is not the only factor that influences the variation of house infestation. Distance between houses has already been shown to influence ecotope reinfestation. It was first demonstrated by Cecere *et al.* (2004), followed by subsequent studies by her and colleagues, leading to the recommendation of vector control interventions to spray all places surrounding an infested spot within a 400-meter radius. Unfortunately, these contributions were not mentioned in the article by Weinberg *et al.* (2019). Trying to show the effect of distance on house infestation by a descriptive analysis looking for a non-random pattern will not automatically encourage health agencies to adopt the use of geospatial tools any better than the previous more elaborated studies.

Weinberg *et al.* (2019) also report that a neural network model to describe house infestation in an area as a post-diction produced a model with a mean accuracy 0.77. There is no elaboration other than a figure, although the conclusion states that the report shows that “it becomes possible to evaluate and plan spray strategies based on predictions of infestation from novel, non-linear models using the spatiotemporal indices implemented”. Neural network model fed with the selected factors is just one of many possible models. Is it better than other possible neural network models? (including other factors), or other models not based on neural networks? Confusing predictions with postdictions, *e.g.*, arguing that novel models are good without a solid base, will not encourage health agencies to adopt geospatial tools within their routine work. The data-driven approach of the machine-learning methods is very good at generating hypothesis that should be tested in subsequent studies, but when the resulting modelling exercise is taken as new knowledge (a well-characterized “questionable research practice”), we can expect a decrease in studies replicability (NAS, 2019), especially in epidemiological systems like Chagas. Machine-learning algorithms are fascinating and suggest sophisticated ways to carry out data analysis. However, important as it is, what are we analysing, what for and for whom take precedence over how are we going to analyse the data.

The “end of theory” advocated by the article written by Anderson (2008), trying to demonstrate that “correlation is enough” (...) because “we can analyze data without hypothesis” (...) throwing “the numbers into the biggest computing clusters the world have ever seen and let statistical algorithms find patterns where science cannot”, was widely and systematically rejected. We should make a strong effort to avoid the attraction of easily calculated correlations of easily available data and search for causal explanations. Discovering a cause is difficult, especially in complex multi-causal systems through observational (*vs* experimental) research, then at least one should make an effort to identify the most parsimonious explanation after considering all the possible alternatives. Thinking that alternative explanations is the key, *i.e.* considering only one alternative when you know there are many, is a mistake.

As anyone familiar with multivariate regression knows, the fraction of “explained” variance adds frequently more than 100% if you carry out regressions for individual independent variables, meaning that the contribution of individual variables depends on the presence and/or interactions with other variables. The case

could be extended to a multivariate regression that will yield significant ($p < 0.05$) coefficients with increasing probability (just by chance) as you increase the number of factors in the regression. One of the best ways to get out of this problem is to let several alternative models compete and let the data be the decider using (for example) AIC values (*i.e.* in a multi-model inference approach) instead of looking for the best model. We must fight the fascination of data dredging using the powerful analytical techniques we have at hand (a battle that I now know I have lost about 20 years back when I started using some of these techniques, *e.g.*, Gorla, 2002).

The article concludes additionally that “We demonstrate a useful and practical method for (“the evaluation of?”) systematic insecticide spraying based on documented previous infestation and spraying. In terms of overall evaluation of the analysed intervention period, most sites presented less infestations, in fact only minor ones, at the latest control included in the study compared to the initial one”. However, besides the suggestion of spraying within 300-meter radius of an infested house (an issue we know from the study by Cecere *et al.*, 2004), it is difficult to identify utility and practicality of the method. The statements “presented less infestations” and “only minor ones” are difficult to sustain because of the mentioned low sensitivity of the man-hour method, and the occurrence of house infestation higher than 5% (high for the vectorial transmission scenarios) in 3 out of 4 localities (according to Figure 2 in the Weinberg’s article) in spite of close surveillance by the FMS personnel, that surely evaluated houses more systematically and regularly than what is usual for state health agencies surveillance.

The statement “In spite of the overall results, we (*sic*) as much as about 33% of persistent infestations despite spraying, possibly due to re-colonization of *T. infestans* from the environment” is rather worrisome because it suggests that after 14 years of close surveillance, the vector control has very low efficacy, meaning that there is either insecticide resistance (not reported so far for the region), there is a problem with the quality of the insecticide application or the integrated program should be revised.

Geospatial tools can make a substantive contribution to support health systems efficacy. However, simply “throwing numbers to the biggest computer cluster” will not convince health systems to adopt the tools for their routine work. Much more effort on the design of the studies is needed to justify the use of the fantastic analytical tools developed during the last decade, and on the clarity of the message to the health agencies as well.

References

- Abad-Franch F, Valenca-Barbosa, Sarquis O, Lima MM, 2014. All that glisters is not gold: sampling-process uncertainty in disease-vector surveys with false-negative and false-positive detections. *PLoS Negl Trop Dis* 8:e3187
- Abad-Franch F, Ferraz G, 2010. “Didn’t you see that bug...?” Investigating disease vector occurrence when detection is imperfect. *Revista Brasileira de Medicina Tropical* 43 (Suppl II):31-4.
- Anderson C, 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*. Available from: <https://www.wired.com/2008/06/pb-theory/#comments>.
- Cavallo MJ, Amelotti I, Abrahan L, Cueto G, Gorla DE, 2018. Rural houses infestation by *Triatoma infestans* in northwestern

- Argentina: Vector control in a high spatial heterogeneous infestation area. *PloS One* 13:e0201391
- Cecere MC, Vazquez-Prokopec GM, Gurtler RE, Kitron U, 2004. Satio-temporal analysis of reinfestation by *Triatoma infestans* (Hemiptera: Reduviidae) following insecticide spraying in a rural community in Northwestern Argentina. *Am J Trop Med Hygiene* Vol 71:803-10.
- Gorla DE, Catalá SS, Grilli MP, 1997. Efecto de la temperatura sobre la distribución de *Triatoma infestans* y el riesgo de transmisión vectorial de la enfermedad de Chagas en Argentina. *Acta Toxicológica Argentina* 5:36-9.
- Gorla DE, 2002. Variables ambientales registradas por sensores remotos como indicadores de la distribución geográfica de *Triatoma infestans* (Heteroptera: Reduviidae). *Ecología Austral* 12:117-27.
- Levy MZ, Malaga-Chavez FS, Cornejo del Carpio JG, Vilhena DA, McKenzie FE, Plotkin JB, 2010. Rational spatio-temporal strategies for controlling a Chagas disease vector in urban environments. *J R Soc Interface*. doi: 10.1098/rsif.2009.0479.
- NAS (National Academies of Sciences, Engineering, and Medicine), 2019. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. ISBN 978-0-309-48616-3. doi:17226/25303.
- Provecho Y, Gaspe MS, Fernandez MP, Gurtler RE, 2017. House Reinfestation with *Triatoma infestans* (Hemiptera: Reduviidae) After Community-Wide Spraying with Insecticides in the Argentine Chaco: A Multifactorial Process. *J Med Entomol* Vol 54:646–57.
- Weinberg D, Lanfri M, Scavuzzo CM, Abril M, Lanfri S, 2019. Evaluation and Planning of Chagas Control Activities Using Geospatial Tools. *Geospat Health* 14:229-38. doi: 10.4081/gh.2019.786.
- Zu Dohna H, Cecere MC, Gurtler RE, Kitron U, Cohen J, 2007. Re-establishment of local populations of vectors of Chagas disease after insecticide spraying. *J Appl Ecol* 44:220-7.
- Zu Dohna H, Cecere MC, Gurtler RE, Kitron U, Cohen J, 2009. Spatial Re-Establishment Dynamics of Local Populations of Vectors of Chagas Disease. *PLOS Negl Trop Dis* 3:e490. doi: 10.1371/journal.pntd.0000490

Non-commercial use only