



A Random Forest-Cellular Automata Modeling Approach to Predict Future Forest Cover Change in Middle Atlas Morocco, Under Anthropic, Biotic and Abiotic Parameters

Anass Legdou¹(✉), Hassan Chafik², Aouatif Amine¹, Said Lahssini³, and Mohamed Berrada²

¹ LGS Laboratory, ENSA Kenitra, Ibn Tofail University, Kenitra, Morocco
anasslegdouseap@gmail.com, aouatif.amine@uit.ac.ma

² Department of Mathematics Informatics, Humanities and Sociology,
University of Moulay Ismail, Meknes, Morocco
chafik1996hassan@gmail.com, berrada.mohamed@gmail.com

³ National Forestry School of Engineers, Sale, Morocco
marghadi@gmail.com

Abstract. This study aims to predict forest species cover changes in the Sidi M'Guild Forest (Mid Atlas, Morocco). Used approach combines remote sensing and GIS and is based on training Cellular Automata and Random Forest (RF) regression model for predicting species cover transition. Five covariates that precludes such transition have been chosen according to Pearson's test. The model was trained and validated based on the use of forest cover stratum transition probabilities between 1990 and 2004 and then validated using 2018 forest species cover map. Validation of the predicted map with that of 2018 shows an overall agreement between the two maps (72%) for each number of RF's trees used. The 2032 projected forest species cover map indicate a strong regression of Cedar atlas and thuriferous juniper cover and a medium regression of mixture holm oak and thuriferous juniper, mixture of atlas cedar and thuriferous juniper, and sylvatic and asylvatic vacuums, a very strong progression of holm oak, and of mixture atlas cedar, holm oak and thuriferous juniper and medium progression of mixture of atlas cedar and holm oak. These findings provide important insights to planners, natural resource managers and policy-makers to reconsider their strategies to ensure the sustainability goals.

Keywords: Forest species cover change · Random forest regression · Cellular automata

1 Introduction

Forests as natural capital provides ecosystem services that contributes to human well being [1]. Sustainable management of forest resources try to maximize the

provided services without compromising their abilities to fulfil these services in the future [2]. Nowadays due to social and human needs, forests resources are under high pressures. In many regions, the pressures are far beyond forests productive capacities. Moreover, climate change increase the fragility of the threatened natural ecosystems. As consequence, a diffuse and still progressing process of land use and forest cover changes are widely documented mainly in developing countries [4]. With the need to maintain forest contribution to global cycles and to protect this natural capital for the next generations, there is a need to understand their dynamic and to predict the tendencies in order to stress urgent actions and formulate appropriate policies to improve land use planning [5]. Such understanding rely on the Land Use Land Change Modeling LULC that try to explain human environment dynamics producing the changes [6]. LULC needs multi temporal land/forest cover maps as well as the driving forces conducting to that changes [7–9]. In addition, machine learning algorithms have been used extensively to explain LULCs. Several researches had combined Cellular automata (CA) with a plethora of modeling frameworks such as Markov chains [10], neural networks [11] support vector machines [12] and kernel-based methods [13] among others. More recently, CA have been successfully combined with Random Forest [14, 15].

Moroccan forests hold a major part of its biological diversity. It covers 5.8 million hectares, including 132,000 ha of cedar, 1.36 million ha of holm oak, 830,000 ha of argan, 350,000 ha of cork oak, 600,000 ha of thuja and 1 million ha of Saharan acacia (HCEFLCD, 1992). The Atlas cedar occupies a prominent place among other species. Moroccan cedar forests, especially those in the Middle Atlas, show regressive trends. Cedar become limited to mountain tops. Among the factors generally blamed for the degradation: regeneration lack due to high grazing pressure [16]; high human pressure (overgrazing, cultivation, illegal cutting, fire, etc.); dieback phenomenon, which is becoming increasingly worrisome about the future of cedar stands, and damages caused by cedar’s natural enemies (defoliating insects, wood-boring insects, fungi and the maggot monkey), which weakens stands stressed by climatic hazards.

Given the status of Moroccan cedar forests as threatened natural capital and in order to understand the driving forces toward its regressive tendencies and to predict its cover in the future, we focus within this work on cedar cover change modeling. The work concerns Sidi M’Guild Forest which belong to Middle Atlas National Park that holds a representative part of Cedar Ecosystem. Land Use change has been modeled using machine learning algorithm predicting future forest cover change as result of driving anthropic, biotic and abiotic factors. Our paper is organized in a Methods describing the used data and algorithms explanation and the achieved results which stress out the main results and the discussions and conclusion.

2 Materials and Methods

The study concerned Sidi M’Guild forest which is located in the Moroccan Middle Atlas. It covers about 29,000 hectares. Cedar covers about 51%, holm oak

34%, and Thuriferous juniper 3.6% of forest area. The overall methodology is shown in the Fig. 1. It consists on getting covariates, then processing and modeling.

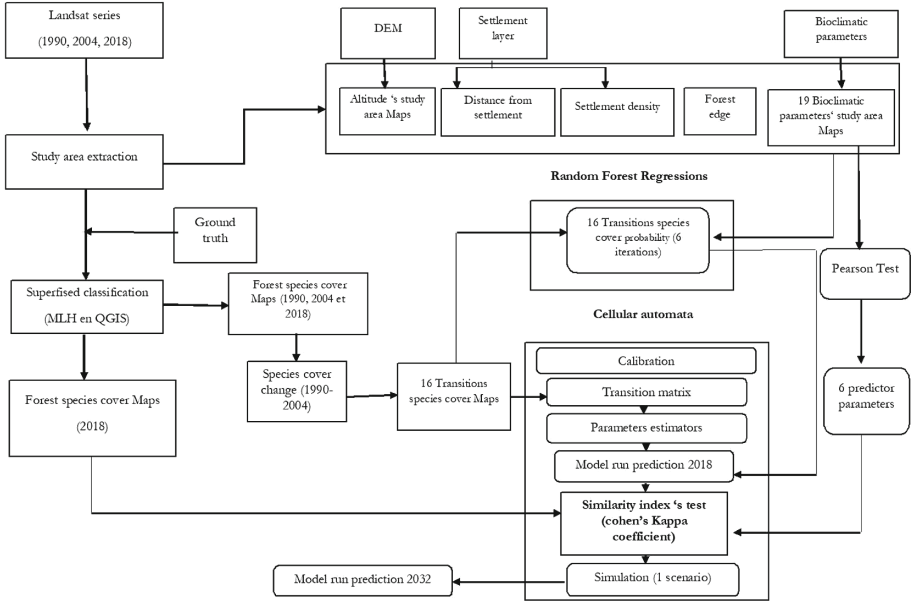


Fig. 1. Schematic of the used approach in this work.

2.1 Used Data

Landsat satellite images covering study used were downloaded from USGS Earth explorer platform. Landsat 4 TM (thematic Mappers) images with a 30 m spatial resolution and 7 spectral bands for 1990 to 2004. Landsat 8 OLI image (Operational Land Image) that contains 11 spectral bands at 30 m spatial resolution were used for 2018. Images were chosen based on the availability and being cloud free and captured during August to reduce atmospheric disturbance and confusion with herbaceous layers' spectral emission. The predictive variables used consists on bioclimatic variables, Digital elevation model (DEM) and human characteristics. Bioclimatic data were downloaded from worldclim's platform (see Table 1)¹. Altitude Maps was extracted from Shuttle Radar Topography Mission (STRM) DEM². The location of human settlement was linked to 2014 Morocco's 'general census' data. Distance from human settlement maps and distance from Forest edge map were generated using basic GIS functions.

Furthermore, a Pearson's r correlation coefficient has been calculated in order to identify the factors that are highly correlated.

¹ <https://www.worldclim.org/>.

² Data collected from <http://dwtkns.com/srtm30m/>.

Table 1. Used bioclimatic variables.

Symbology	Designation
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp–min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5–BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

2.2 Preprocessing and Image Classification

Satellite data were preprocessed QGIS and then classified using Maximum Likelihood algorithm. Images were standardized and accommodated to the same extent and spatial resolution. Image classification was based on training the classifier using, as ground truth, forest stand type maps released in 1990, 2004 and 2018. Six dominant forest cover stratum/classes were chosen while respecting the National forest inventory standards. The resulting classes are: atlas cedar (Ca), holm oak (Qr), thuriferous juniper (Jt), atlas cedar and thuriferous juniper mixture (CaJt), atlas cedar and holm oak mixture (CaQr), mixture of atlas cedar, holm oak and thuriferous juniper (CaQrJt), holm oak and thuriferous juniper mixture (QrJt) and sylvatic and asylvatic vacuums (V). In addition to Atlas cedar, we classified other species closed to or mixed with Atlas cedar, since when we talk about the cedar ecosystem, we refer to all the species that are closely linked to it. The accuracy assessment based on cross-validation of the classified images showed an overall accuracy of 88.86%, 90.21% and 92.33% respectively for the years 1990, 2004 and 2018.

2.3 Random Forest Regression and Calibration Model

Random Forest is a flexible and easy to use machine learning algorithm which performs both regression and classification tasks [17]. It combines multiple decision trees in determining the final output. It uses different technics for training the model. Bagging technique [18], which involves training each decision tree on a random set of data sampled, without replacement, from the training data set with a random split selection [19]. Random forest regression, which the performance has been proved in several studies [20,21], was used to predict forest cover class transition probabilities. Transition maps for the period 1990–2004 were established by observing the behavior of the forest species. For each transition from one class to another, two possible values 1 denotes change from class to other class, and 0 denotes no change. Then, binary transition maps were produced 0 for stability and 1 for change occurrence and used for training 16 models. Following the approach adopted by Gounaridis et al. [15], the transition probability surfaces were generated through training Random Forest algorithm [18] using all variables and using the most independent covariates identified through the Pearson’s correlation coefficient calculation (5 factors). The RF regression models were then implemented in python using the Random Forest regression. The model output is transition probability map. The model was run for each of the sixteen transitions defined before (Fig. 2) and for a set on number of trees (10, 20, 30, 40, 50, 100).

Cellular automata [22] were used to predict the future state of forest species distribution. Probability maps obtained with different number of trees were produced and for each time, the Kappa coefficient of Cohen was calculated through comparing predicted and observed results of the same year [23].

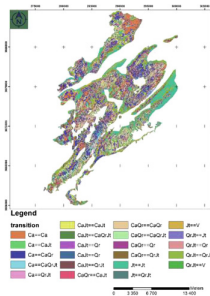


Fig. 2. Forest species cover change, during 1990–2000.

3 Results and Discussion

3.1 Forest Species Cover Classification

Accuracy assessment based on cross-validation showed an overall accuracy of 88.86%, 90.21% and 92.33% respectively for the years 1990, 2004 and 2018.

Thus, indicating the suitability of the derived classified maps for effective and reliable forest species cover change analysis and modeling. Post-classification analysis of the spatial metrics and their variations based on Table 2 showed that the area occupied by Ca, CaQr, CaQrJt and Jt classes had drastically decreased between 1990 and 2018 in the studied area. On the other hand, the area occupied by Qr, CaJt, QrJt and V had substantially increased during the same period. Such results seems to be coherent with literature describing Cedar as vulnerable species and holm oak as green cement with a high adaptive capacities.

Table 2. Temporal distribution of forest species cover classes (ha) and percentages of change.

Classes	Area size (ha)			Change in forest species cover ($\Delta\%$)		
	1990	2004	2018	2004–1990	2018–2004	2018–1990
Ca	8975,88	2171,34	2007,85	-313	-8	-347
CaQr	8482,05	9141,93	7138,62	7	-28	-19
Qr	44,37	551,88	233,1	92	-137	81
CaQrJt	4034,52	3184,83	3758,58	-27	15	-7
Jt	5010,75	3491,28	4854,94	-44	28	-3
QrJt	2207,97	5986,71	4802,95	63	-25	54
CaJt	2207,97	2405,7	2282,89	8	-5	3
V	0,99	1822,86	3677,6	100	50	100

3.2 Analysis of Transition Probabilities and Model Validation

Using all factors as covariates or retaining the five factors that are not correlated according to Pearson’s r correlation coefficient (Temperature Seasonality: Bio4, precipitation of Driest Quarter: Bio17, Distance from human settlement, settlement density and distance from Forest edge) gave the same results (Table 3). Such fact could be explained by RF robustness toward correlated data. In general, we notice that the higher scores were recorded for numbers of trees equal to 50 and 100 trees. In addition, we conclude that the five non correlated parameters considerably affects the evolution of the forest species change between 1990 and 2004. As explained in the methods, the predictive model was used to predict 2018 forest cover state. The model returns simulated maps of forest species cover distribution for the year 2018, relative to each of RF model number of trees. The predicted maps were compared to existing 2018 forest cover maps and the validation was based on the use of Kappa coefficient. Validation results are given in Table 4.

We notice that there was no evident difference between the values achieved by Kappa Coefficient. Hence, we could conclude that the number of trees used in the random forest model does not influence the results of the simulation. Contrariwise, the thresholds experimented for each rule did influence the model.

Table 3. Transition probability of forest species cover areas for each number of trees.

Transition of species between 1990 and 2004	Area size	Training score					
		n = 10	n = 20	n = 30	n = 40	n = 50	n = 100
Ca-CaJt	223,56	0,79	0,79	0,8	0,8	0,8	0,81
Ca-QrJt	121,68	0,86	0,86	0,87	0,87	0,88	0,88
Ca-CaQrJt	509,13	0,82	0,82	0,83	0,84	0,84	0,84
Ca-CaQr	5945,58	0,42	0,42	0,45	0,46	0,47	0,48
CaQr-CaJt	1658,25	0,53	0,53	0,55	0,56	0,57	0,58
CaQr-QrJt	987,21	0,76	0,76	0,77	0,78	0,78	0,79
CaQr-CaQrJt	2459,7	0,32	0,32	0,35	0,36	0,37	0,38
CaQr-Qr	167,22	0,59	0,59	0,61	0,62	0,62	0,63
Jt-V	992,25	0,45	0,45	0,48	0,48	0,49	0,5
Jt-QrJt	166,77	0,45	0,45	0,48	0,49	0,49	0,5
QrJt-V	827,01	0,38	0,38	0,41	0,42	0,43	0,45
QrJt-Jt	589,23	0,45	0,45	0,47	0,49	0,41	0,51
QrJt-Qr	162,81	0,54	0,54	0,56	0,57	0,57	0,58
CaJt-QrJt	1338,66	0,24	0,24	0,27	0,28	0,29	0,31
CaJt-CaQrJt	188,28	0,24	0,24	0,27	0,29	0,3	0,32
CaJt-Qr	191,61	0,58	0,58	0,6	0,61	0,61	0,62

Table 4. Kappa’s coefficient for each RF number of trees

Trees number	n = 10	n = 20	n = 30	n = 40	n = 50	n = 100
Kappa de Cohen	0.7213	0.72126	0.72127	0.72125	0.7212	0.7211

3.3 Model Comparison and Predictions for 2034

By comparing the areas occupied by each class in both simulated and observed maps, we observe some differences (Table 5):

Table 5. Comparison of the actual and simulated forest cover species in 2018

Classes	Observed 2018 (ha)	Simulated 2018 (ha)	Difference simulated observed (ha)	Simulated observed in 2018 ($\Delta\%$)
Ca	2007.85	1703.07	-304.78	-17,90
CaQr	7138.62	6513.03	-625.59	-9,61
Qr	233.1	1473.03	+1239.93	84,18
CaQrJt	3758.58	3707.1	-51.48	-1,39
Jt	4854.94	5069.34	+214.4	4,23
QrJt	4802.95	4768.92	-34.03	-0,71
CaJt	2282.89	2305.71	+22.82	0,99
V	3677.6	3200.67	-476.93	-14,90

The model did closely estimate the area of CaQr, CaQrJt, Jt, QrJt and CaJt. On the other hand, it did largely overestimated the area of Qr about -18% and miss estimated the area occupied by Ca, CaQr and V. In addition, the

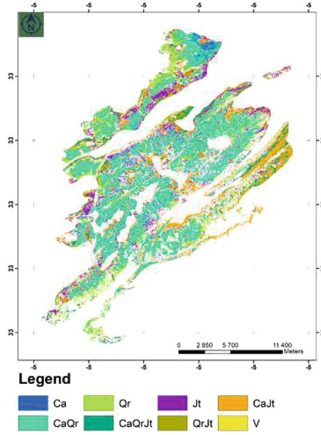


Fig. 3. Forest species cover predicted for 2032.

Table 6. Comparison of forest cover species change between 2018 and 2032

Class	Observed 2018 (ha)	Simulated 2032	Rate change (%)
Ca	2007.85	170.82	-91.49
CaQr	7138.62	9144.99	+28.11
Qr	233.1	3249.45	+1294.02
CaQrJt	3758.58	8444.61	+124.68
Jt	4854.94	1243.62	-74.39
QrJt	4802.95	2405.52	-49.92
CaJt	2282.89	1254.51	-45.05
V	3677.6	2881.44	-21.65

model has predicted a regression of Ca, CaQr, CaQrJt, QrJt and V and then the area affected by the change were localized. It predicted Qr, Jt and CaJt progression. The model overestimated some species classes and underestimated others because we did not take into account the interventions carried out by the administration in terms of silvicultural interventions on holm oak and cedar stands, the reforestation actions undertaken including silvicultural vacuums, the awareness campaigns conducted for the benefit of the populations bordering the forest and finally the forest police actions to preserve and conserve the forest heritage. Theses interventions must be deeply analyzed in order to assess their impacts on explaining the occurred forest cover changes.

With regard to the predictions, the model returned the forest species cover state for 2032 (Fig. 3 and Table 6).

The model predicted a strong regression for Ca and Jt cover and a medium regression for QrJt, CaJt, and V. It did also predicate a very strong progression for Qr and CaQrJt and medium progression for CaQr. The majority of forest

species cover change are depending on their positioning to settlements and to forest's boundaries. Stratum that will show regression trends are located near human settlements and not far from forest edge. On the other hand, stratum with progressive trends are located far from human settlements and to forest's boundaries. Such finding seems to be coherent with literature and forest managers declarations.

4 Conclusion

The current study was based on an integrated approach that combines remote sensing and GIS to simulate and predict plausible forest species cover changes for Sidi M'Guild Forest for the years 2032 using Cellular Automata (CA)-Random Forest regression (RF) model. The initial forest species cover map (1990), the transition potential maps (1990–2004) and the 1990–2004 transition probabilities were used to train RF model. Model was validated using actual and predicted 2018 forest species cover. The overall agreement between the two maps was 72% for each number of RF's trees used. The future 2032 projections indicate a strong regression of Cedar atlas with -91.49% and thuriferous juniper cover with -74.39% and a medium regression of holm oak and thuriferous juniper mixture with -49.92% , atlas cedar and thuriferous juniper mixture with -45.05% , and sylvatic and asylvatic vacuums with 21.65% , a very strong progression of holm oak with $+1294.02\%$, and of atlas cedar, holm oak and thuriferous juniper mixture with $+124.68\%$ and medium progression of atlas cedar and holm oak mixture with $+28.11\%$ by 2032. The majority of forest species cover changes depends on their location to settlements and to forest's boundaries. Regression are located near human settlement and forest boundaries. As cedar is considered as national heritage, these findings could be useful for decision makers and for managers to review their strategies in order to ensure the sustainability of cedar as natural capital.

References

1. Costanza, R., et al.: Quality of life: an approach integrating opportunities, human needs, and subjective well-being. *Ecol. Econ.* **61**(2–3), 267–276 (2007)
2. Bettinger, P., Boston, K., Siry, J., Grebner, D.: *Forest Management and Planning*, p. 362. Academic Press, Cambridge (2017)
3. Bettinger, P., Chung, W.: The key literature of, and trends in, forest-level management planning in North America, 1950–2001. *Int. For. Rev.* **6**, 40–50 (2004)
4. Giri, C.P., Pengra, B.W., Long, J.B., Loveland, T.R.: Next generation of global land cover characterization, mapping and monitoring. *Int. J. Appl. Earth Obs. Geoinf.* **25**, 30–37 (2013)
5. Tena, T.M., Mwaanga, P., Nguvulu, A.: Impact of land use/land cover change on hydrological components in chongwe river catchment. *Sustainability* **11**, 6415 (2019)
6. Mertens, B., Lambin, E.: Modelling land cover dynamics: integration of fine-scale land cover data with landscape attributes. *Int. J. Appl. Earth Obs. Geoinf.* **1**, 48–52 (1999)

7. Yang, F., Liu, Y., Xu, L., Li, K., Hu, P., Chen, J.: Vegetation-ice-bare land cover conversion in the oceanic glacial region of tibet based on multiple machine learning classifications. *Remote Sens.* **12**(6), 999 (2020)
8. Abdi, A.M.: Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GISci. Remote Sens.* **57**, 1–20 (2019)
9. Akubia, J.E.K., Bruns, A.: Unravelling the frontiers of urban growth: spatio-temporal dynamics of land- use change and urban expansion in Greater Accra Metropolitan Area, Ghana. *Land* **8**(9), 131 (2019)
10. Fu, X., Wang, X.Y., Yang, J.: Deriving suitability factors for CA-Markov land use simulation model based on local historical data. *J. Environ. Manag.* **206**, 10–19 (2018)
11. Islama, K., Rahmanb, M.F., Jashimuddinc, M.: Modeling land use change using Cellular Automata and Artificial Neural Network: the case of Chunati Wildlife Sanctuary, Bangladesh. *Ecol. Ind.* **88**, 439–453 (2018)
12. Mustafa, A., Rienow, A., Saadi, I., Cools, M., Teller, J.: Comparing support vector machines with logistic regression for calibrating cellular automata land use change models. *Eur. J. Remote Sens.* **51**, 391–401 (2018)
13. Liu, X., et al.: A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landsc. Urban Plan.* **168**, 94–116 (2017)
14. Gounaridis, D., Chorianopoulos, I., Symeonakis, E., Koukoulas, S.: A Random Forest-Cellular Automata modelling approach to explore future land use/cover change in Attica (Greece), under different socio-economic realities and scales. *Sci. Total Environ.* **646**, 320–335 (2019)
15. Gounaridis, D., Chorianopoulos, I., Koukoulas, S.: Exploring prospective urban growth trends under different economic outlooks and land-use planning scenarios: the case of Athens. *Appl. Geogr.* **90**, 134–144 (2018)
16. Moukrim, S., et al.: *Agrofor. Syst.* **93**(4), 1209–1219 (2019)
17. Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: *Classification and Regression Trees*. Chapman and Hall/CRC, Boca Raton (1984)
18. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1996). <https://doi.org/10.1007/BF00058655>
19. Dietterich, T.: An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting and randomization. *Mach. Learn.* **32**, 1–22 (1998)
20. Zhao, Q., Yu, S., Zhao, F., Tian, L., Zhao, Z.: Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. *Forest Ecol. Manag.* **434**, 224–234 (2019)
21. Srinet, R., Nandy, S., Patel, N.R.: Estimating leaf area index and light extinction coefficient using Random Forest regression algorithm in a tropical moist deciduous forest, India. *Ecol. Inform.* **52**, 94–102 (2019)
22. Kamusoko, C., Gamba, J.: Simulating urban growth using a Random Forest-Cellular Automata (RF-CA) model. *Int. J. Geo-Inf.* **4**(2), 447–470 (2015)
23. Zapf, A., Castell, S., Morawietz, L., Karch, A.: Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate? *BMC Med. Res. Methodol.* **16**, 93 (2016). <https://doi.org/10.1186/s12874-016-0200-9>