SAGA Application for Generalized Estimating Equations Analysis

Luís Moncaixa[1,2(✉)] and Ana Cristina Braga[1,2]

[1] School of Engineering, University of Minho, Braga, Portugal
luismoncaixa@hotmail.com
[2] ALGORITMI Research Centre, LASI, University of Minho, Campus de Gualtar, 4710-57 Braga, Portugal

Abstract. Logistic regression models seek to identify the influence of different variables/factors on a response variable of interest. These are normally used in the field of medicine as it allows verifying which factors influence the presence of certain pathologies. However, most of these models do not consider the correlation between the variables under study. In order to overcome this problem, GEE (Generalized Estimating Equations) models were developed, which consider the existing correlation in the data, resulting in a more rigorous analysis of the influence of different factors. There are different packages in R that allow analysis using GEE models, however, their use requires some prior knowledge of the R programming language. In order to fill this gap and enable any user to perform analysis through GEE models, a Shiny application called SAGA (**S**hiny **A**pplication for **G**EE **A**nalysis) was developed. The developed web application is available for use at the following link http://geemodelapp2022.shinyapps.io/Shiny_App. The main purpose of the SAGA application is to develop and analyse GEE models using a dataset selected by the user, where it will be possible to describe all the variables of interest in the development of the model, as well as validate the same models developed through validation by ROC analysis. In addition to the results of the GEE models, shown in the application, the ROC curves of each developed model are also represented.

Keywords: Logistic Regression · Correlated Data · GEE · Shiny · SAGA

1 Introduction

The use of statistical methods and associated research in human investigation, especially in the field of medicine, has demonstrated in recent decades to be a fundamental element for the development of treatments for diseases as well as for scientific research [3].

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14105, pp. 53–68, 2023.
https://doi.org/10.1007/978-3-031-37108-0_4

Longitudinal data indicate repeated observations over time for the same sample/individual, with observations more identical within the same individual than across different individuals, thus indicating that these observations will be correlated [5].

This type of data, given the correlation among observations, represents an additional difficulty when applying regression models for analysis, since one of the main assumptions of these models is based on the independence of observations, that is, the correlation is non-existent. Several biomedical analyses of longitudinal data demonstrate that the use of regression models disregarding the correlation of observations results in a less precise estimation of standard errors, influencing all indicators such as confidence intervals or p-values, which allow for a good analysis and can lead to incorrect inferences [8].

The GEE models (Generalized Estimating Equations), developed by *Liang and Zeger* [9], represent a class of models that are often used for data in which the responses are correlated. These models are an extension of GLMs (Generalized Linear Models), as they take into account the correlation present in the data, allowing their application to longitudinal data.

Unlike GLMs, which are based on maximum likelihood theory, the structure of GEE models is related to quasi-likelihood theory, so it does not require specifying the distribution of the data, only requiring the specification of the relationship between the response variable and the covariates, as well as the relationship between the mean and variance of the responses through a correlation structure [2,18]. The inclusion of this correlation structure distinguishes GEE models from any conventional logistic regression model, however, it is not necessary to specify this same structure since GEE models can estimate consistent parameters without any specified correlation structure, as demonstrated by *Liang and Zeger* [9].

Although the inclusion of the correlation structure is not necessary in the development of GEE models, the presence of a structure appropriate to the data being studied allows for better parameter estimates, resulting in more valid statistical inferences. There are several types of correlation structures with the most commonly used being the independence, exchangeable, autoregressive and unstructured. The independent correlation structure indicates that the repeated observations within an individual are independent, meaning they are not correlated. The exchangeable correlation structure assumes that any two or more responses within the same cluster have the same correlation, and the order of the responses is arbitrary. The autoregressive correlation structure commonly applied to studies where time dependence is assumed, meaning that the correlation between responses depends on the time interval between them, with responses with shorter time being more correlated. Finally, the unstructured correlation structure represents different correlations between each observation [21,23].

According to *Horton and Lipsitz* [6], each correlation structure presented earlier should be selected depending on the sample being analyzed. However, there are situations where it is difficult to identify the correlation structure that best fits the sample. To assist in the selection of the correlation structure, *Pan* [12] developed a method called quasi-likelihood under the independence model

criterion, QIC. This method is an adaptation of the Akaike information criterion method (AIC), used to estimate the quality of different models, allowing for the selection of the best model, used in GLM models. Since the AIC method is based on the properties of maximum likelihood and these do not apply to GEE models, the QIC method was developed based on the quasi-likelihood method developed by *Wedderburn* [19], thus allowing for its applicability in GEE models.

For different GEE models with different correlation structures, several QIC values are generated, so the best model and, in turn, the best correlation structure will be the model with the lowest QIC value. If different correlation structures present equal QIC values, one should consider the QICu value, which represents a simplified version of QIC [12].

Several authors use GEE models to identify the influences of different factors in biomedical research. *Mitchell et al.* [10] sought to verify if pseudoexfoliation would influence risk factors for the development of glaucoma. Through logistic regression models, they found that the presence of ocular pseudoexfoliation is related to glaucoma in older individuals. However, through GEE models, they found that the risk factors associated with glaucoma are independent of the presence of pseudoexfoliation.

The Shiny package enables the creation of dynamic applications that simplify the use and analysis of GEE models for biomedical studies by allowing users to apply these models without prior programming or code-typing experience in order to derive statistical inferences that allow for the quick and accurate performance of analyses through statistical tests or graphical displays [1].

The objective of this article is to present the new application developed for GEE analysis, SAGA, detailing all the sections and menus existing in it, applied to a case study. Initially, the methodology and approaches to be used in the development of the application will be described, as well as a description of the software and packages that will be used. The case study used in the demonstration of the application will also be introduced. Next, all the menus of the developed application will be demonstrated in detail, ending with a brief conclusion and future prospects.

2 Material and Methods

Statistical programs for GEE models are available in different statistical analysis software products such as *Stata*, *R*, *SPSS*, among others. To verify limitations between different packages, *Nooraee et al.* [11] used a set of packages available in different languages, including *GENMOD*, *SAS*, *GENLIN*, *SPSS*, `repolr`, `multgee`, and `geepack` in R. This author found that the *multgee* and `repolr` packages have more functionalities than the others, but all of them showed limitations in parameter estimation when the number of individuals is very low.

Several packages are available in R for GEE analysis, the `geepack` [22] package was selected for implementation in the Shiny SAGA application, as this package allows for parameter estimation using the GEE methodology, generating statistical inferences and making it possible to use the models developed for

validation. Implemented together with the `geepack` package is the `ROCR` package [17], which allows for validation of the developed models, enabling analysis between models, and complementing this analysis with graphical representation of the generated ROC curves.

2.1 Case of Study

To perform the GEE analysis a dataset related to the identification of the presence of ocular pseudoexfoliation in several patients undergoing cataract removal surgery during June 1 to December 31 of 2016 was used, this dataset was provided by the Department of Ophthalmology of the Vila Nova de Gaia/Espinho Hospital Center situated in Vila Nova de Gaia, Portugal.

In addition to personal information about the patients such as age, sex, and date of consultation, over time, information was collected on the eye under analysis, presence of pseudoexfoliation, pupil dilation, presence of intraocular lens and associated diopters, as well as the location where it is inserted.

Pseudoexfoliation syndrome (PEX) is an age-related pathology characterized by fibrillar material deposits, usually identified by a white color, present in the ocular structures of the anterior segment bathed by the aqueous humor [13,15]. This disease is associated with abnormal behaviors and changes in the LOXL1 gene, responsible for synthesizing the fibrous material present in ocular pseudoexfoliation, as reported by *Sangal and Chen* [14], *Govetto et al.* [4] and *Jammal et al.* [7]. Despite this association, there are also external factors associated with ocular pseudoexfoliation, such as age, sex, and geographical area of residence, as reported by *Sangal and Chen* [14] and *Shazly et al.* [16].

2.2 Shiny

The development of web applications for statistical analysis presented itself as a challenge, as it required some deep knowledge of other languages, such as JavaScript, HTML or CSS to develop the application interface and establish a connection with the R language for statistical processing and analysis [20].

Shiny is an R package that enables the development of web applications without requiring prior knowledge of other programming languages, using only the R language. The application developed in Shiny is composed of two components, the interface, which represents the component that corresponds to the structure of the application responsible for receiving all the instructions provided by the user, which will be used to perform the desired tasks in order to visualize the results obtained, and the server, which represents the component responsible for all the processing of the instructions provided by the user in order to perform the desired statistical analysis through functions/algorithms developed in R.

Shiny uses reactive programming, which allows updating in real-time all the results and processes performed, when a change occurs in the instructions provided by the user. After developing the two essential components, described previously, the application is available on the web for any user for free.

Although there are several packages in different programming languages for the application of GEE model analysis, there are not actually web applications

that allow the use of these GEE analysis in an intuitive way without programming knowledge. The SAGA application seeks to solve the presented gap, making it possible through it:

- Load the relevant dataset for analysis;
- Select the type of variables, correlation structure, as well as edit the data;
- Validate the developed models and compare them;
- Analyze the obtained results intuitively;
- Export the generated graphical representations;
- Obtain help in the application's functioning.

To describe all the steps to be performed by the user during the use of SAGA for GEE analysis, a flowchart was constructed, as presented in Fig. 1.

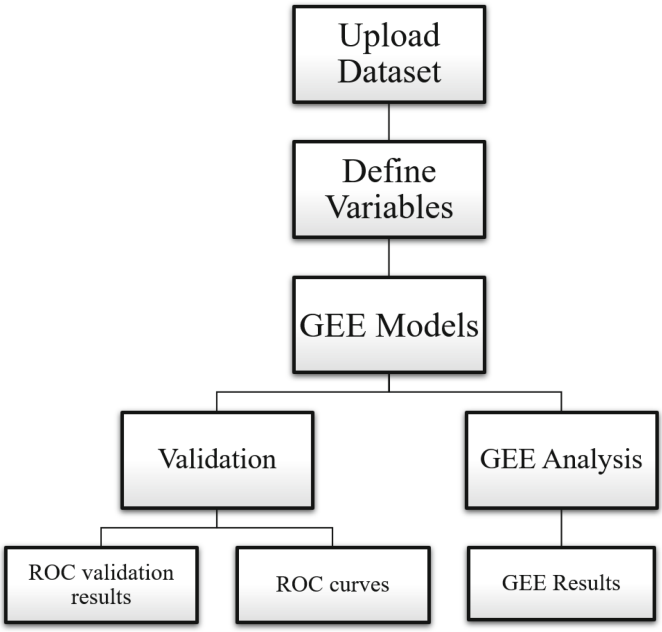


Fig. 1. Flowchart of the structure of the Shiny SAGA application.

Initially, the user must load the dataset to be used, then select all the variables of interest for the construction of the GEE model. Subsequently, the user can view the results of the GEE analysis or validate the created model through receiver operating characteristic validation, where the results from the respective validation as well as the generated ROC curves are obtained.

3 Results

In this section, all the menus and sections existing in the developed SAGA application will be described in detail. The language used will be English in order to facilitate the use among various users.

The application is divided into three sub-menus as described in Fig. 2, which are:

A. Upload and editing of the dataset;
B. Selection of variables for the GEE model;
C. Validation and results of the generated models.

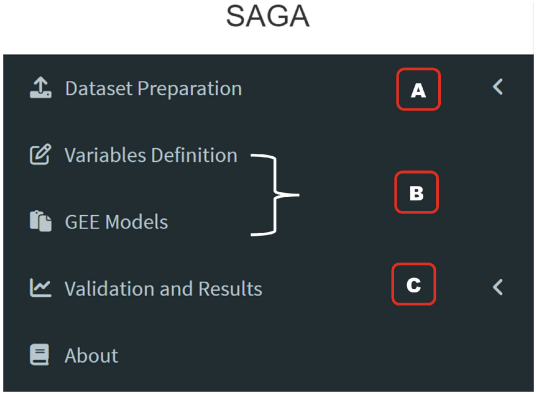


Fig. 2. SAGA menu.

Throughout this section, all menus and submenus of the SAGA application will be described in detail using the dataset corresponding to the case study as an example.

3.1 SAGA: Upload and Editing the Dataset

This menu is based on the user's introduction of the dataset and, if necessary, making changes to it for further development of GEE models, and it is divided into two submenus, upload dataset and dataset changes.

SAGA: Upload Dataset. In this menu, the user will introduce the dataset by clicking the "browse..." button, a new window will appear to select the dataset, limited only to EXCEL files with the `.xlsx` extension and CSV files with .csv extension. Any other data structure will generate an error when imported. The dataset will appear in the background window. In addition to downloading the dataset, there are also some features that can improve or change the visualization of the dataset, as represented in Fig. 3.

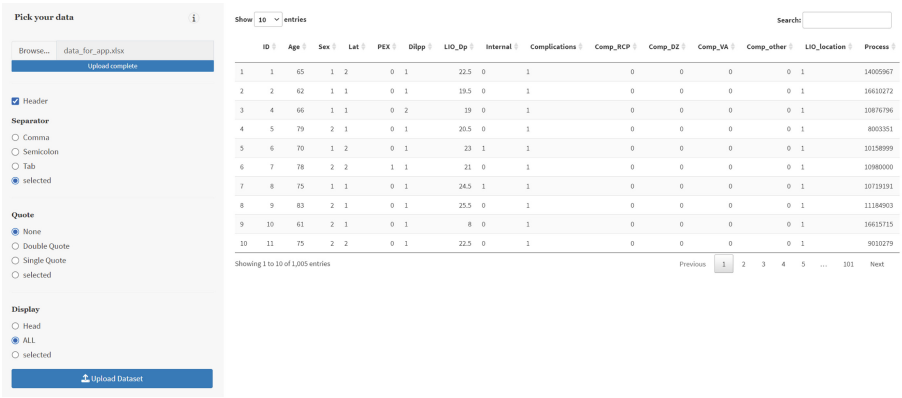


Fig. 3. Importing the dataset as .xlsx file.

SAGA: Data Changes. After the upload of the dataset, there may be some changes that can be made, for example the name of the respective existing columns. In the case of GEE models, column names containing spaces between names should be avoided, as these names often generate errors when the model analysis is performed. The change of the column names can be made in the "Columns Names" section, where all the names of the columns existing in the previously downloaded dataset are displayed, and a text box where the new column name will be entered. To make the change, the desired column will be selected, and its new name will be entered in the text box, and the change will be changed automatically. An example of the use of this section is shown in Fig. 4.

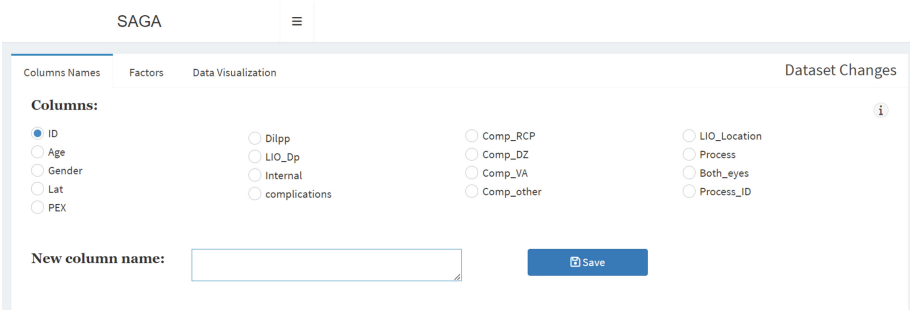


Fig. 4. Column name change from Data Changes menu.

In addition to changing the column names of the dataset, the user can factorize columns whose values are categorical variables. With factorization, these values will be divided into levels, one level for each corresponding variable in the selected column, making it easier to use in the development of GEE models. This process can be done on the sub-menu "Factors". The last sub-menu will show again the uploaded dataset with all the changes made by the user.

3.2 SAGA: Selection of Variables for GEE Models

As described earlier, the variable selection menu corresponds to the second part of the application, and is one of the most fundamental, as all the variables necessary for the development of GEE models will be defined here. The variables that will be selected are:

- Response variable;
- Covariates;
- Identification column, represents the column corresponding to the sample/observation identification;
- Distribution Family, represents the type of distribution of the response variable (e.g. Binomial, Poisson, etc.);
- Correlation Structure.

SAGA: Variables. The selection of variables begins with the selection of the response variable of the GEE model and its corresponding covariates. In the selection of the response variable, all columns corresponding to the dataset uploaded previously are present, and the column corresponding to the variable of interest will be selected. The selection of covariates and their relationship can be defined in two ways:

- **Selection of covariates:** Represented as **A** in Fig. 5. Similar to the response variable, all columns of the dataset will be presented, and the user can select the columns he/she wants. Through this selection process, the covariates will be added to each other. (Example: Age + Gender... + Latency)

- **Introduction of the relationship between covariates:** Represented as **B** in Fig. 5. Another method for selecting covariates is by introducing them and their relationship to each other. The user must input the names of the columns of interest and specify the relationship between the selected covariates.

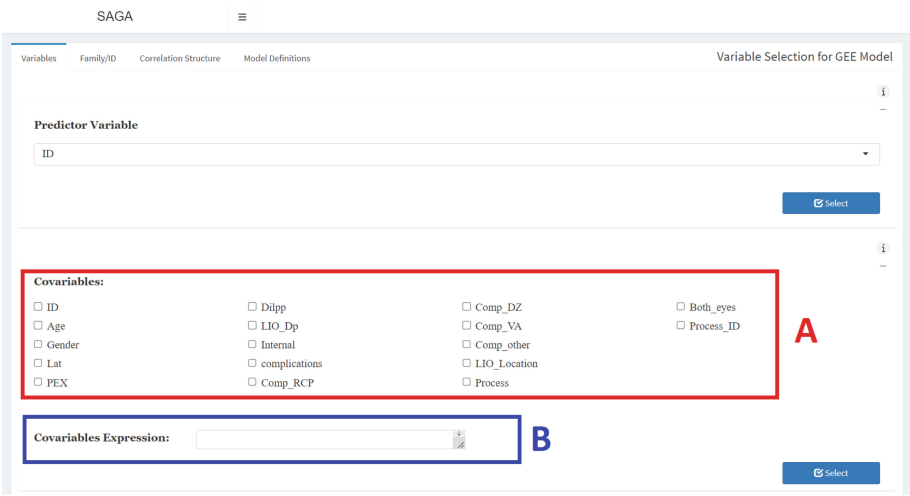


Fig. 5. Selection of the response variable and the process of selecting covariates.

The user can only choose one of the processes for selecting covariates. If both processes are used, the process **A** will be assumed.

SAGA: Family/ID. In this sub-menu, the distribution family of the response variable and the column corresponding to the sample/observation identification are defined. The process for selecting these two variables is the same as described for the response variable, where all the columns of the dataset will be presented, and the user should choose the column of interest.

SAGA: Correlation Structure. The selection of the correlation structure will be the last step in developing the GEE model. The user can select one of the three available correlation structures, described before, which are:

- Independent correlation structure;
- Exchangeable correlation structure;
- Autoregressive correlation structure;

If the user is unsure which correlation structure best fits the model, there is an option to display QIC values by selecting the "Display QIC values" option. This option will present a table with the corresponding QIC and QICu values for each correlation structure, considering the variables previously selected, as represented in Fig. 6. If any variable from the previous sub-menus are missing, this option will deliver an error.

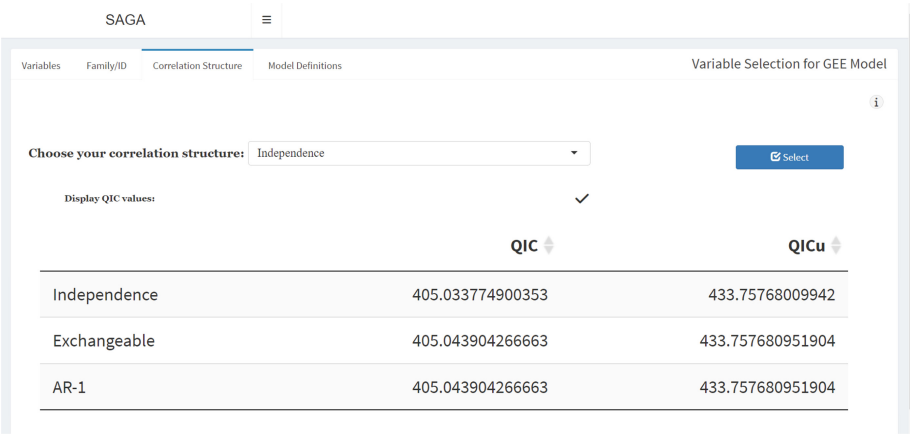


Fig. 6. Selection of the correlation structure with QIC values option.

For the case study, based on the results shown regarding the correlation structure, it is observed that the QIC values are very similar in all structures. However, the independent correlation structure has a slightly lower value. Therefore, the selected structure for developing the GEE models will be the independent correlation structure.

SAGA: Model Definitions. In this final sub-menu, a name will be assigned to the newly developed GEE model and a brief description of the model. This section will not influence the developed model, it is only necessary to label the generated results.

3.3 SAGA: GEE Models

In this menu, the GEE models created are presented, describing the previously selected variables, an example of this menu is shown in Fig. 7. There are 3 functionalities in this menu:

– **Add:** This functionality allows the construction of a new GEE model in addition to the first created model. When the button is executed, the user will return to the variable selection process in the Model Definitions submenu, and should select "Variables" as the initial section and continue the procedure described above. The new model will be added to the menu when the "Select" button in the "Model definitions" sub-menu is executed. There is a limit of 3 models created, and the button will become inactive after creating that number of models.

- **Edit:** To execute this button, one of the created models must be selected, and the button will be inactive when two or more models are selected. When this button is executed, the user can change the variables selected in the previous procedure. The model change occurs when the "Save" button in the "Model definitions" sub-menu is executed, so it is suggested to check all sections corresponding to the previously described variables.
- **Run:** The execution of this button occurs when one or more models are selected. Through this button, the variables selected by the user will be considered, and there is no possibility of changing them. The considered variables allow the GEE analysis to be performed, as well as their use in the ROC validation phase, which will be described next.

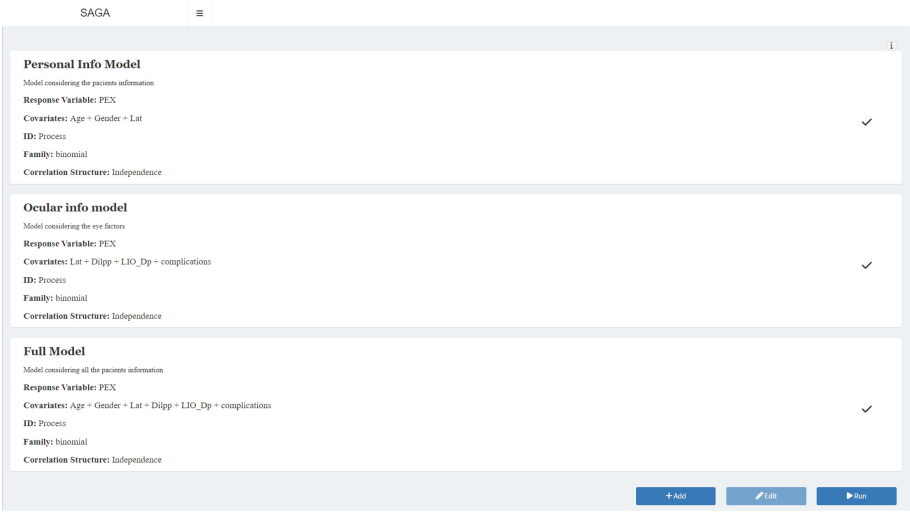


Fig. 7. Representation of the GEE models menu with three models as an example.

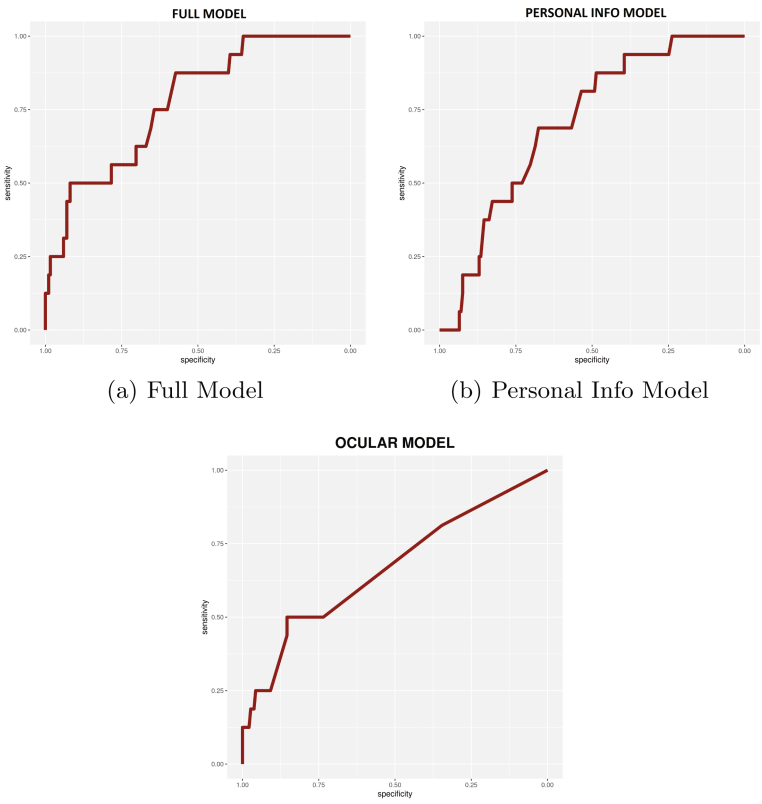
As shown in Fig. 7, for the case study, three different models were developed: one considering only the personal variables of the patients, another considering the ocular variables, and finally a model considering all variables together.

3.4 Validation: ROC Analysis

The validation through ROC analysis of the developed models is an additional and optional functionality, that means, that the user is not required to perform the validation to access the results of the GEE analysis. However, through this validation, the user can obtain additional information about the model or compare them through the metrics obtained during the analysis. Through the ROC validation, the ROC curves will be generated and the value of two metrics,

accuracy and area under the curve. These two metrics will help to check the performance of the model. This validation procedure can be done in the "Validation and Results" menu.

Data Split. The first step to proceed with the validation of the models is to define the division of the dataset into training and testing. At this stage, the user can select three different divisions, and in case they are not sure which division to use, there is a default selection that divides the dataset considering its size. After choosing the desired division for validation, the user must press the button "Run ROC" to perform the validation.



(a) Full Model (b) Personal Info Model

(c) Ocular Model

Fig. 8. Generated ROC curves for the three different models.

ROC: Graph Visualization. After initializing the ROC analysis in the previous sub-menu, the ROC curves generated for each of the models selected in the previous sections are presented. Below each of the generated curves, there is a "download" button where the user can download the obtained curves to their local computer. The file is in `.png` format. Figure 8 shows the respective ROC curves considering three different models and their respective download buttons.

ROC: Results. As a complement to the graphical representation of ROC curves generated for different models, this section presents the results obtained in the ROC analysis considering the two metrics referenced in previous sections, Accuracy, it shows how accurate a collection of measurements are in relation to their true value, and AUC. For each previously selected model, the values of the described metrics will be presented, allowing the user to verify, through the obtained results, the model that presents better validation results. Figure 9 show how the results from ROC validation are presented for each developed model.

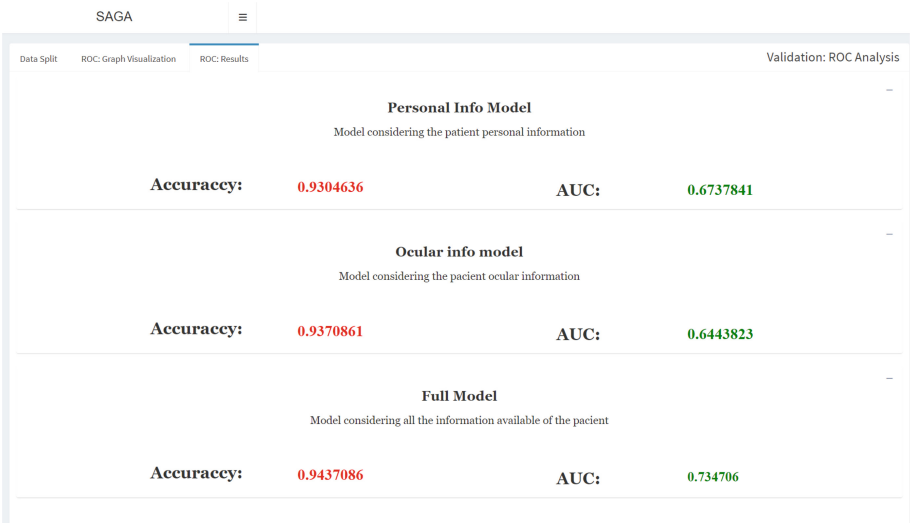
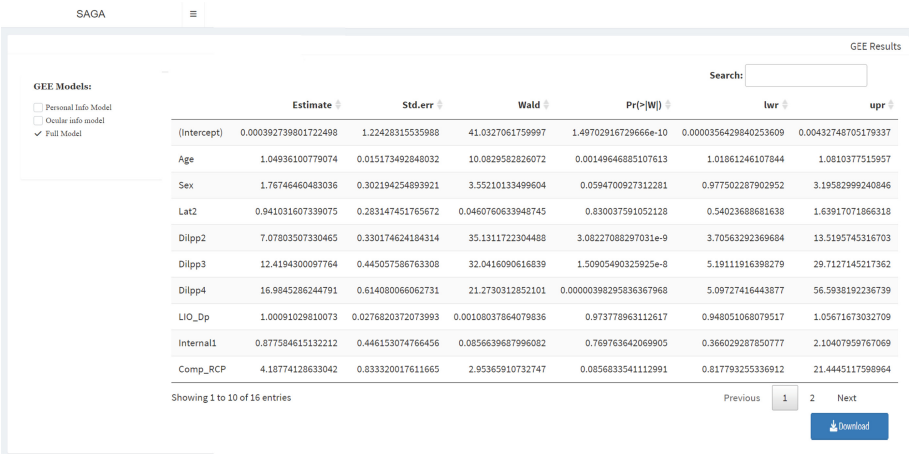


Fig. 9. Example of results from ROC validation for different models.

Based on the ROC validation, it is possible to verify that the model considering all variables presents higher values in both metrics (ACC and AUC) compared to the other models. Therefore, this will be the model considered for the GEE analysis of the case study.

3.5 GEE Results

The results obtained from the models created in the previous menus are presented in this menu. For each developed model, a table describing the results obtained from the GEE analysis will be presented. The obtained table allows the user to analyze the influence of covariates on the response variable, through their estimate values, confidence interval values and the Wald test. These results enable the user to perform a rigorous analysis of the developed model throughout all the steps. The results' table can be downloaded to the user's personal computer in the `.csv` format, allowing for download through the download button for further analysis or visualization. Figure 10 shows a table with the results obtained from one of the models created during the description of all previous menus.



	Estimate	Std.err	Wald	Pr(>|W|)	lwr	upr
(Intercept)	0.000392739801722498	1.22428315535988	41.0327061759997	1.497029167296666e-10	0.000035629840253609	0.00432748705179337
Age	1.049361007790 74	0.015173492848032	10.0829582826072	0.00149646885107613	1.01861246107844	1.0810377515957
Sex	1.76746460483036	0.302194254893921	3.55210133499604	0.0594700927312281	0.977502287902952	3.19582999240846
Lat2	0.941031607339075	0.283147451765672	0.0460760633948745	0.830037591052128	0.54023688681638	1.63917071866318
Dilpp2	7.07803507330465	0.330174624184314	35.1311722304488	3.082270882970 31e-9	3.70563292369684	13.5195745316703
Dilpp3	12.4194300097764	0.445057586763308	32.0416090616839	1.509054903259 25e-8	5.19111916398279	29.7127145217362
Dilpp4	16.9845286244791	0.614080066062731	21.2730312852101	0.000003982958 36367968	5.09727416443877	56.5938192236739
LIO_Dp	1.00091029810073	0.0276820372073993	0.00180037864079836	0.973778963112617	0.948051068079517	1.05671673032709
Internal1	0.877584615132212	0.446153074766456	0.0856639687996082	0.769763642069905	0.366029287850777	2.10407959767069
Comp_RCP	4.18774128633042	0.833320017611665	2.95365910732747	0.0856833541112991	0.817793255336912	21.4445117598964

Fig. 10. Results obtained for one of the models created in the previous menus.

In general, it can be observed through GEE analysis using the SAGA application, that the covariates corresponding to age, pupillary dilation and two complications that appear during cataract removal surgery, have a positive influence on the presence of ocular pseudoexfoliation.

The validation of this influence is based on the obtained p-values lower than 0.05, the estimates values for the influential covariates are positive, and the confidence intervals, which have values greater than 1, demonstrate the covariates' significance.

As already discussed by other authors and demonstrated in the results, age and pupil dilation continue to have an influence on the presence or development of ocular pseudoexfoliation. Also, through the results, it is observed that complications during cataract removal surgery, observed when a zonular

dehiscence occurs, may indicate the presence of ocular pseudoexfoliation, as infections resulting from the registered dehiscence may arise, including pseudoexfoliation itself.

In Table 1, all results obtained from the GEE analysis using the SAGA application applied to the case study are described.

Table 1. Results of GEE analysis for the case study

	B	**SE(B)**	W	$Pr(>\|W\|)$	CI - 95%
Intercept	0.000399	1.253	39.005	4.23e–10	[3.42e-05–0.005]
Age	1.049	0.015	10.700	0.001	[1.020–1.080]
Gender	1.769	0.301	3.582	0.058	[0.980–3.192]
Lat (LE)	0.941	0.283	0.046	0.830	[0.540–1.638]
LIO DP	1.000	0.028	1.09e-03	0.974	[0.948–1.060]
LIO Location (Sulcus)	2.152	0.786	0.952	0.329	[0.460–10.040]
LIO Location (CA)	9.65e+17	1.346	946.875	0.0	[6.90e+16–1.35e19]
LIO Location (S/LIO)	4.60e–17	1.943	374.704	0.0	[1.012e-18–2.07e-15]
Dilpp (≥7–8)	7.078	0.330	35.142	3.07e–09	[3.706–13.519]
Dilpp (≥6–7)	12.414	0.444	32.211	1.38e–08	[5.20–29.63]
Dilpp (≥5–6)	16.999	0.614	21.272	3.99e–06	[5.10–56.62]
Internal (S)	0.877	0.446	0.086	0.769	[0.37–2.10]
Comp RCP (S)	4.188	0.833	2.954	0.086	[0.82–21.45]
Comp DZ (S)	10.769	0.803	8.767	0.003	[2.23–51.93]
Comp VA (S)	0.365	1.373	0.540	0.463	[0.02–5.38]
Comp other (S)	14.718	0.748	12.916	0.0003	[3.40–63.79]

B: Parameter estimate; SE(B): Standard Error; W: Wald statistic; CI 95%: 95% confidence interval.

4 Conclusions and Future Work

To address the lack of an application exclusively designed for GEE analysis of longitudinal data, SAGA was developed and completed. The application allows for the development of different GEE models considering various scenarios, performs validation through ROC curves for each model, generates the obtained curves and their respective metrics, enabling a more rigorous analysis of the developed models. Finally, it describes the results of the GEE analysis for each model, allowing users to save the obtained results on their personal computer throughout the various processes. This application allows for a more interactive and faster GEE analysis by the user.

This article represents the first release of the SAGA application. The application can suffer bug fixes and improvements on the visual component. SAGA requires the input dataset to be pre-processed, meaning that there are not many tools in the application that allow for data pre-processing before constructing GEE models, such as removing missing data. The introduction of a limitation to access to all the menus before select the variables, can enable a better analysis and reduce the likelihood of errors or undefined parameters.

References

1. Chang, W., et al.: shiny: Web Application Framework for R (2023). https://shiny.rstudio.com/, r package version 1.7.4.9002
2. Christopher, Z.: Generalized estimating equation models for correlated data: a review with applications. Am. J. Polit. Sci. **45**(2), 470–490 (2001)
3. Fitzmaurice, G.M., Laird, N.M., Ware, J.H.: Applied Longitudinal Analysis, vol. 998. John Wiley & Sons (2012)
4. Govetto, A., et al.: Frequency of pseudo exfoliation among patients scheduled for cataract surgery. J. Cataract Refract. Surg. **41**(6), 1224–1231 (2015)
5. Hedeker, D., Gibbons, R.D.: Longitudinal Data Analysis. Wiley-Interscience (2006)
6. Horton, N.J., Lipsitz, S.R.: Review of software to fit generalized estimating equation regression models. Am. Stat. **53**(2), 160–169 (1999)
7. Jammal, H., et al.: Characteristics of patients with pseudoexfoliation syndrome at a tertiary eye care center in jordan: a retrospective chart review. Ophthalmol. Therapy **10**(1), 51–61 (2021)
8. Kleinbaum, D.G., Klein, M.: Logistic regresion for correlated data: Gee. Logist. Regression Self-learn. Text 327–375 (2002)
9. Liang, K.Y., Zeger, S.L.: Longitudinal data analysis using generalized linear models. Biometrika **73**(1), 13–22 (1986)
10. Mitchell, P., Wang, J.J., Hourihan, F.: The relationship between glaucoma and pseudoexfoliation: the blue mountains eye study. Arch. Ophthalmol. **117**(10), 1319–1324 (1999)
11. Nooraee, N., Molenberghs, G., van den Heuvel, E.R.: Gee for longitudinal ordinal data: comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN. Comput. Stat. Data Analy. **77**, 70–83 (2014)
12. Pan, W.: Akaike's information criterion in generalized estimating equations. Biometrics **57**(1), 120–125 (2001)
13. Ritch, R., Schlötzer-Schrehardt, U.: Exfoliation syndrome. Surv. Ophthalmol. **45**(4), 265–315 (2001)
14. Sangal, N., Chen, T.C.: Cataract surgery in pseudoexfoliation syndrome. In: Seminars in Ophthalmology, vol. 29, pp. 403–408. Taylor & Francis (2014)
15. Schlötzer-Schrehardt, U., Naumann, G.O.: Ocular and systemic pseudoexfoliation syndrome. Am. J. Ophthalmol. **141**(5), 921–937 (2006)
16. Shazly, T.A., Farrag, A.N., Kamel, A., Al-Hussaini, A.K.: Prevalence of pseudoexfoliation syndrome and pseudoexfoliation glaucoma in upper Egypt. BMC Ophthalmol. **11**(1), 1–6 (2011)
17. Sing, T., Sander, O., Beerenwinkel, N., Lengauer, T.: Package 'ROCR'. Visualizing the performance of scoring classifiers, 1–14 (2015)
18. Wang, Y.G., Fu, L., Paul, S.: Analysis of Longitudinal Data with Examples (2022)
19. Wedderburn, R.W.: Quasi-likelihood functions, generalized linear models, and the gauss-newton method. Biometrika **61**(3), 439–447 (1974)
20. Wickham, H.: Mastering Shiny. O'Reilly Media, Inc. (2021)
21. Wilson, J.R., Lorenz, K.A.: Modeling Binary Correlated Responses using SAS, SPSS and R, vol. 9. Springer (2015)
22. Yan, J., Højsgaard, S., Yan, M.J., False, B.: Package 'geepack' (2012)
23. Ziegler, A., Vens, M.: Generalized estimating equations. Meth. Inf. Med. **49**(05), 421–425 (2010)