



HAL
open science

Biostatistiques médicales avec GMRC Shiny Stats - un outil de formation par la pratique

Thibaut Fabacher, Michael Schaeffer, Nicolas Tuzin, Francois Severac,
François Lefebvre, Marie Mielcarek, Eric-Andre Sauleau, Nicolas Meyer,
Julien Godet

► **To cite this version:**

Thibaut Fabacher, Michael Schaeffer, Nicolas Tuzin, Francois Severac, François Lefebvre, et al.. Biostatistiques médicales avec GMRC Shiny Stats - un outil de formation par la pratique. *Annales Pharmaceutiques Françaises*, 2020, pp.499-506. 10.1016/j.pharma.2020.06.001 . hal-03020600

HAL Id: hal-03020600

<https://hal.science/hal-03020600>

Submitted on 23 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Biostatistiques médicales avec *GMRC Shiny Stats* - un outil de formation par la pratique

Thibaut Fabacher¹, Michael Schaeffer², Nicolas Tuzin¹, François Séverac¹, François Lefebvre¹, Marie Mielcarek¹, Erik-André Sauleau¹, Nicolas Meyer¹, and Julien Godet^{1,✉}

¹Groupe Méthode Recherche Clinique - Hôpitaux Universitaires de Strasbourg, 1 place de l'hôpital, 67000 Strasbourg, France

²Société Alstats, 3 route de Kientzville, 67750 Scherwiller, France

Les biostatistiques sont omniprésentes dans la littérature scientifique et médicale et constituent une compétence indispensable à tout étudiant en santé. Nous avons développé un outil de formation par la pratique aux biostatistiques.

***GMRC Shiny Stats* est une application interactive spécifiquement dédiée aux analyses statistiques de données médicales. L'application permet de réaliser une succession d'analyses qui mime la démarche usuelle d'un statisticien expérimenté face à un jeu de données. *GMRC Shiny Stats* permet de réaliser les analyses les plus courantes (analyses descriptives, inférences selon des méthodes fréquentistes, comparaison des caractéristiques de populations, analyses de survie, corrélation, concordance, etc.). La facilité d'utilisation et l'accompagnement de l'utilisateur dans le choix des tests statistiques constituent les intérêts principaux de l'outil. Ces fonctionnalités permettent aux étudiants de se former en réalisant des analyses statistiques de manière autonome dans le cadre de travaux universitaires.**

Biostatistiques | Pédagogie | Analyse de données
Correspondence: julien.godet@chru-strasbourg.fr

Introduction

Les biostatistiques sont omniprésentes dans la littérature scientifique et médicale ainsi que dans les rapports d'évaluation et de recommandation émanant des agences de santé. Les étudiants en santé (médecine, pharmacie, odontologie, maïeutique,...) se doivent d'acquérir des connaissances et des compétences en biostatistiques qui leur permettent de comprendre et analyser de façon critique ces sources documentaires (1, 2). Au cours de leur formation, les étudiants bénéficient d'unités d'enseignements (en général de premier cycle) dédiées à l'acquisition de ces compétences, le plus souvent sous la forme de cours traditionnels d'introduction aux biostatistiques. Mais ces enseignements sont souvent considérés par les étudiants comme étant essentiellement mathématiques et leur pertinence pour leur future activité professionnelle n'apparaît pas toujours évidente (3).

Plus tardivement dans leur cursus, les étudiants sont amenés à réaliser des travaux d'initiation à la recherche. Très souvent il s'agit d'études (études cliniques rétrospectives, études d'évaluation des pratiques,...) qui requièrent de constituer un recueil de données et de procéder à son analyse. Les étudiants sont donc confrontés à la nécessité de réaliser des analyses statistiques plus ou moins complexes. Leurs en-

cadranrs leur recommandent souvent de faire appel à un statisticien pour les guider dans cette tâche (4). Ce dernier peut être amené à réaliser tout ou partie du travail d'analyse des données, de sorte que les étapes d'analyse entre le recueil des données et la rédaction des résultats peuvent demeurer mal maîtrisées en raison d'une implication limitée de l'apprenant dans ces tâches. L'informatisation massive des données médicales au cours de ces dernières années a conduit à augmenter le volume des données analysables. Le temps de collecte de ces données a quant à lui très largement diminué et il devient en effet beaucoup moins fastidieux d'extraire des informations fiables des dossiers médicaux informatisés ou de questionnaires en ligne. Le temps économisé à l'étape de la collecte peut donc en partie être reporté vers le travail d'analyse statistique. Dans cette optique, la mise à disposition d'outils faciles à prendre en main peut permettre aux étudiants de participer activement à cette étape en réalisant par eux même un certain nombre d'analyses.

Nous avons développé *GMRC Shiny Stats*, une application interactive et d'utilisation intuitive qui est spécifiquement dédiée aux analyses de données médicales dans le cadre de travaux universitaires (thèses, mémoires, ...). *GMRC Shiny Stats* permet de réaliser des analyses descriptives complètes ainsi que des analyses graphiques de l'ensemble des variables, qu'elles soient de nature qualitatives ou quantitatives. L'inférence statistique est réalisée par des méthodes fréquentistes sur la base de tests d'hypothèse nulle de comparaison de groupes en terme de moyennes, de distributions, proportions ou de survie. Il est également possible de réaliser des mesures de corrélation ou des mesures de concordance. Enfin, il est possible de déterminer les performances diagnostiques de tests. L'ensemble de ces fonctionnalités sont accompagnées de fichiers d'aide et d'un ensemble de vidéos tutorielles qui sont disponibles en ligne pour accompagner les étudiants dans la réalisation et la compréhension de leurs travaux d'analyses statistiques.

Matériel et Méthodes

GMRC Shiny Stats a été développé aux Hôpitaux Universitaires de Strasbourg par le Groupe Méthode en Recherche Clinique (GMRC). *GMRC Shiny Stats* est une application interactive basée sur le logiciel et langage de statistiques

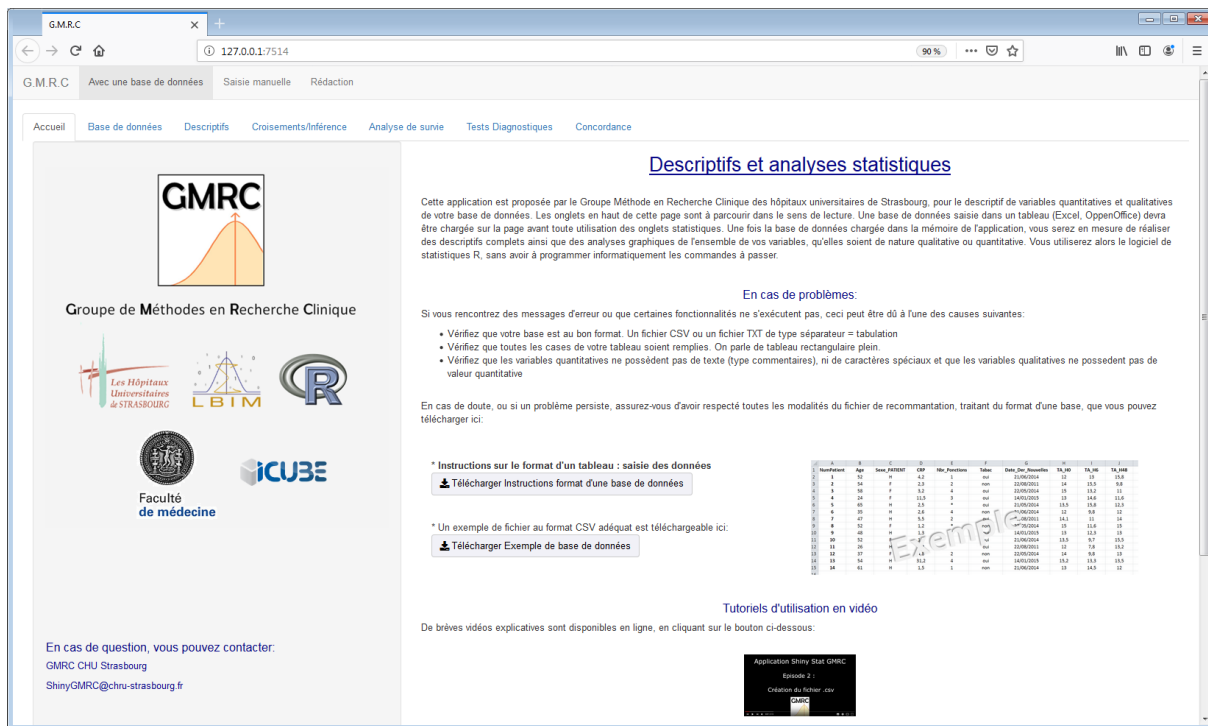


Fig. 1. Aperçu de la page d'accueil de GMRC Shiny Stats.

L'application s'exécute dans un navigateur internet. Un système d'onglets permet d'accéder aux pages dédiées aux différentes étapes d'une analyse statistique de données

R (<https://cran.r-project.org>) construite à partir du package shiny (5).

L'application se présente sous forme d'une page html (Figure 1) qui s'ouvre dans n'importe quel navigateur internet récent. R s'exécute en arrière-plan mais cette opération est totalement transparente pour l'utilisateur.

Plusieurs solutions techniques sont proposées pour la distribution de l'application. Aux Hôpitaux Universitaires de Strasbourg, l'application est disponible pour l'ensemble des personnels (internes, cliniciens, etc.) ayant accès au serveur intranet du CHRU.

Une version portable est également disponible sous forme d'une application autonome sans fichier d'installation. Celle-ci embarque une version portable de R et du navigateur Google Chrome permettant de travailler sur une session locale d'une machine non connectée à partir d'une simple clé USB par exemple. Une version MacOS et une version Windows ont été développées et sont téléchargeables gratuitement. Une version en développement (nightly build) - éventuellement moins stable mais qui inclut les dernières contributions - est également disponible. Enfin, pour ceux qui disposent déjà d'un ordinateur sur lequel R et Rstudio sont installés, l'application existe également sous forme d'un package R. Ce dernier peut s'installer directement à partir de GitHub - ce qui rend l'application plate-forme indépendante et permet donc également son utilisation sous environnement Linux. Le code source est disponible sur GitHub sous <https://github.com/jgodet/GmrcShinyStats>. Contrairement aux versions stables dont les nouvelles versions sont rendues disponibles annuellement ou bi-annuellement, le package est évolutif et les améliorations proposées par des développeurs

de la communauté peuvent être ajoutées régulièrement. L'application est ouverte, libre (licence GPLv3) et gratuite.

Résultats

GMRC Shiny stats, une application spécifiquement dédiée aux analyses de données médicales notamment dans le cadre des travaux universitaires. C'est une application interactive particulièrement simple à prendre en main.

L'application a été pensée pour proposer un workflow d'analyse qui correspond à la démarche usuelle d'un statisticien face à un jeu de données à analyser - à savoir chargement des données ; description des données manquantes puis des variables du jeu de données (permettant l'inspection des données pour au besoin nettoyer ou recoder des variables) ; enfin modélisation des données et éventuellement inférence à partir de l'échantillonnage analysé.

Chargement des données.

Les bases de données saisies dans un tableur de type Excel ou LibreOffice sous format de type texte (*.txt, *.csv) peuvent être importées directement dans l'application. La visualisation de la base permet de vérifier que le chargement des données s'est bien déroulé. Cela permet également de vérifier que les données ont été préalablement bien préparées (saisie homogène, recodage, changement d'échelle, etc.).

Analyse descriptive.

La description des données est généralement la première étape de l'exploration des données (6). La description statistique comprend des statistiques sur les tendances centrales

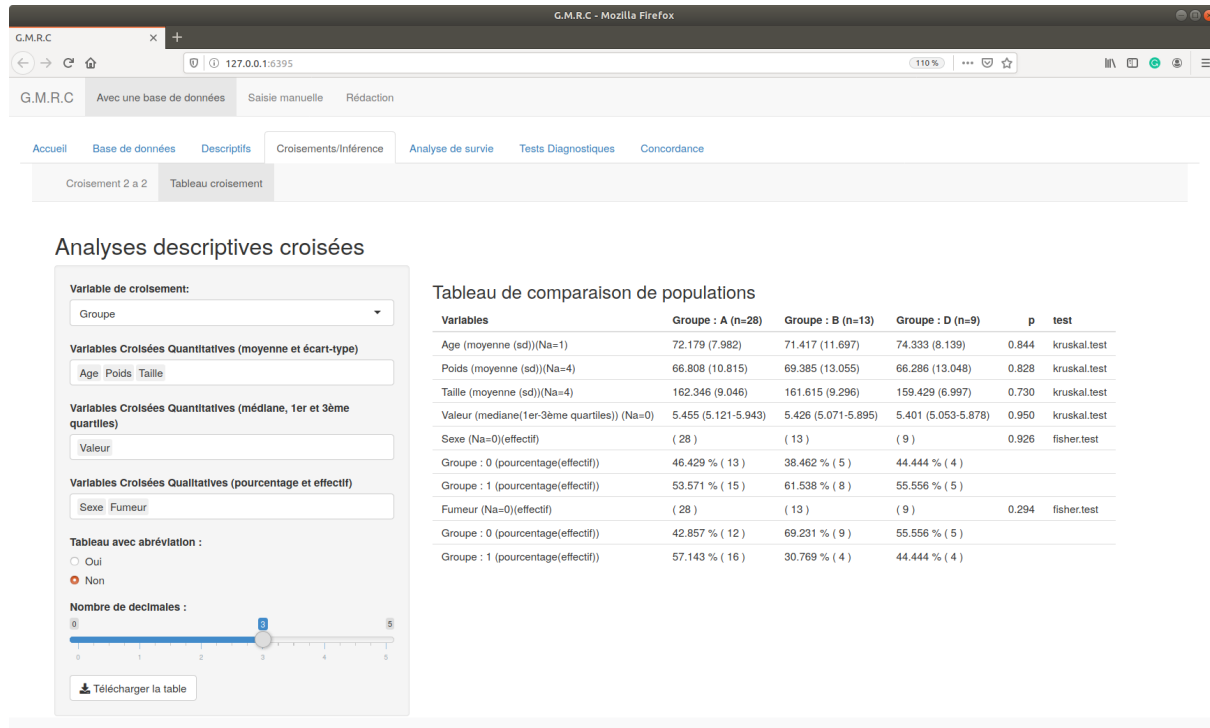


Fig. 2. Aperçu de l'onglet permettant de construire le tableau des caractéristiques des patients de l'étude par groupe.

telles que la moyenne et la médiane ; ainsi que sur la dispersion des variables avec des indicateurs comme l'écart-type ou l'intervalle inter-quartile.

Une étape critique, pourtant souvent négligée, de l'analyse descriptive consiste - avant même de décrire les variables d'une base de données - à explorer les données manquantes du recueil (7). Celles-ci peuvent être sources de biais et il est important de décrire avec soin les données manquantes pour pouvoir interpréter correctement les résultats qui seront obtenus. Ceci est proposé dans le premier sous-onglet dans lequel les données manquantes du jeu de données sont décrites et représentées graphiquement. Cette étape permet également de pointer d'éventuels manquements à l'étape de saisie et compléter, si c'est envisageable, les données.

Le second sous-onglet permet à l'utilisateur de réaliser les analyses descriptives univariées des variables. L'utilisateur sélectionne dans un menu déroulant la variable à décrire et précise si celle-ci est une variable quantitative ou qualitative. Le descriptif complet correspondant à la nature de la variable ainsi que des représentations graphiques de la variable sont alors proposés. L'exploration descriptive attentive des données permet également de repérer les données extrêmes, possiblement aberrantes, erronées ou mal renseignées (mauvais encodage) appelant le cas échéant à un nouveau nettoyage des données.

Inférence et croisement de variables.

L'inférence désigne les techniques statistiques qui permettent

d'utiliser les observations faites sur un échantillon pour les extrapoler à une population. *GMRC Shiny Stats* permet de réaliser de l'inférence selon des approches fréquentistes dont l'interprétation repose en partie sur la notion de p-valeur.

Un certain nombre de tests courants sont disponibles dont le test t de Student, le test de Mann-Whitney, le test du χ^2 , le test exact de Fischer, l'analyse de la variance (paramétrique et non-paramétrique), les tests de corrélation (Pearson, Spearman), des tests de normalité, etc. La problématique du choix du test pour comparer des groupes est souvent source d'incertitudes et d'erreurs (8-11). L'application *GMRC Shiny Stats* guide l'utilisateur dans ce choix. En effet, ce dernier définit uniquement la nature des variables (quantitative ou qualitative) à comparer et un algorithme propose et réalise le test qui semble le plus approprié pour décrire les données. Le choix du test à réaliser répond à une algorithmique qui se base sur la nature des variables à comparer et sur des résultats de tests préalables réalisés automatiquement visant à vérifier que les conditions d'applications du test (normalité, homoscedasticité, etc.) sont réunies.

Un second onglet (intitulé tableau croisement) (Figure 2) permet de construire ce qui constitue souvent le tableau 1 dans les publications de recherche biomédicale, c'est-à-dire le tableau des caractéristiques des patients de l'étude décrites par groupe (par exemple par bras de traitement). Il suffit à l'utilisateur de définir la variable qui définit les groupes, puis de choisir les variables des caractéristiques à faire apparaître dans le tableau. À noter que pour les variables quantitatives,

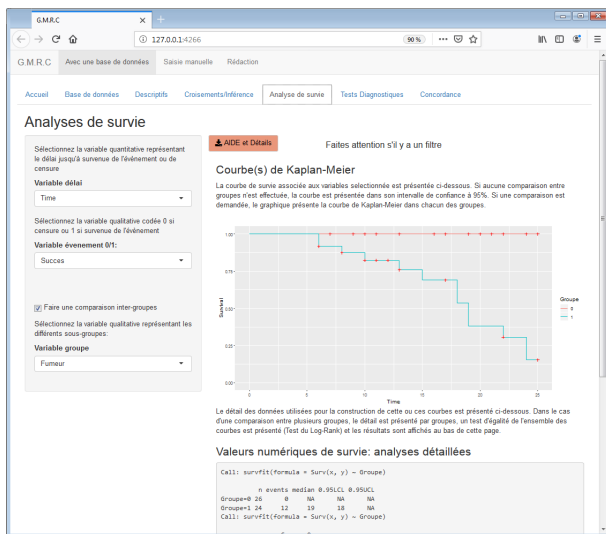


Fig. 3. Aperçu de l'onglet permettant de réaliser des analyses de survie.

il est possible de les décrire soit par leur moyenne et écart-type (si la variable est considérée comme distribuée normalement – information donnée dans le sous onglet Croisement 2 à 2), sinon par leur médiane et intervalle inter-quartile. La p-valeur et le type de test utilisé sont également mentionnés. Ce tableau peut-être téléchargé sous forme csv pour faciliter la préparation des rapports d'analyses ou des publications.

Survie.

L'analyse de survie permet l'étude de données telles que des temps d'observation jusqu'à la survenue d'un événement. Dans ces analyses, tous les sujets ne présentent pas nécessairement l'évènement d'intérêt (on parle de sujets censurés). Pour ces derniers, il est possible de définir une durée d'observation correspondant à la période de suivi. Ainsi, l'analyse de survie correspond à l'analyse conjointe de ces deux informations qui sont encodées par deux variables distinctes : une variable délai qui précise la durée d'observation et une variable évènement qui indique la survenue ou non de l'évènement (Figure 3). Dans *GMRC Shiny Stats*, les courbes de survie sont construites selon l'estimation non-paramétrique de Kaplan-Meier avec leurs intervalles de confiance associés (package Survival (12)). Afin de comparer les survies entre différents groupes, l'utilisateur saisit en plus dans un menu déroulant la variable permettant de définir les groupes à comparer. Le test du log-rank (13) ou sa généralisation à plus de deux groupes sont utilisés pour comparer les courbes de survie.

Évaluation des performances diagnostiques.

La valeur informationnelle d'un test diagnostique correspond à son aptitude à discriminer le statut de sujets vis-à-vis d'une pathologie (malade ou non-malade). Dans *GMRC Shiny Stats*, plusieurs indicateurs sont calculés pour étudier la performance d'un test. La courbe ROC permet de déterminer le meilleur seuil de dichotomie de la mesure quantitative sur la variable à prédire et le pouvoir discriminant du test (mesure de l'aire sous la courbe ROC ou AUC – package pROC (14)). Les nombres de vrais positifs (VP), de vrais

négatifs (VN), de faux positifs (FP) et de faux négatifs (FN) sont déterminés. Les paramètres de sensibilité, de spécificité, les valeurs prédictives et un taux d'exactitude sont calculés.

Mesure de concordance.

L'analyse de l'accord entre observateurs constitue souvent un moyen utile pour évaluer la fiabilité d'un système de cotation. Des mesures concordantes indiqueraient l'existence d'un consensus dans le diagnostic et suggère que les mesures d'intérêt sont reproductibles. Le coefficient kappa (κ) est un indice pour mesurer l'accord entre des valeurs binaires ou nominales entre deux (15) ou plus de deux (16) évaluateurs. *GMRC Shiny Stats* propose uniquement le calcul du κ de Cohen (package irr). Une table d'ordre de grandeur des valeurs de κ (17) est proposée pour aider à l'interprétation de ce paramètre mais celle-ci reste délicate notamment parce que le nombre de catégories influe sur l'estimation obtenue (18).

Cas des analyses en sous-groupes.

Les analyses en sous-groupes ont pour but d'explorer un effet dans une sous-population de patients d'une étude. Il s'agit souvent d'évaluations post hoc. Ceci peut être réalisé facilement dans *GMRC Shiny Stats* en ajoutant un/des filtre(s) sur la base de donnée dans l'onglet « Base de donnée ». Les filtres se combinent à l'aide de l'opérateur logique ET (AND en anglais) et conservent toutes les lignes du tableau de données qui répondent à l'ensemble des conditions sélectionnées. L'ensemble des analyses descriptives ou inférentielles réalisées dans les onglets suivants ne portent alors plus que sur le sous-groupe sélectionné. Il est important de rappeler que les résultats des analyses en sous-groupe ne doivent pas être envisagés comme preuve de démonstration (19, 20). Au contraire, ces analyses doivent rester à vocation exploratoires puisque l'analyse en sous-groupes vise à générer des hypothèses pour des recherches futures. Les résultats obtenus peuvent être alors considérés comme un préalable à d'autres études cliniques. En ce sens il reste intéressant dans un cadre réflexif de pouvoir réaliser ces explorations facilement.

Cas des données appariées.

Les données appariées sont des données qui ne sont pas indépendantes. L'absence d'indépendance résulte souvent de la répétition d'une même mesure sur un même individu. L'illustration classique est celle de mesures faites avant et après l'application d'un traitement chez un malade. La plupart des tests proposés par *GMRC Shiny Stats* supposent l'indépendance des données. Pour ne pas créer de confusion, le choix a volontairement été fait de ne pas ajouter la possibilité de choisir entre données indépendantes et appariées lors de la réalisation des tests. Dans la plupart des cas, il est quand même possible de réaliser des tests de données quantitatives appariées. Pour cela, l'utilisateur doit faire lui-même la démarche d'encoder une nouvelle variable qui correspond à des différences intra-sujets (par exemple une nouvelle variable qui vaudrait la différence entre les valeurs avant et après traitement). Il nous a semblé que cette approche permet une meilleure prise de conscience de ce que sont les données ap-

pariées, de la démarche qui peut être utilisée pour les rendre indépendantes et donc de l'interprétation des résultats qui peut en être faite.

Discussion et Conclusions

Si la problématique de l'accessibilité des outils informatiques pour les analyses statistiques a pu être limitant dans le passé, il existe maintenant un certain nombre d'excellents logiciels ou applications qui permettent d'effectuer de nombreux tests statistiques dans un temps très court et qui sont disponibles gratuitement. On peut citer par exemple le projet BiostaTGV <https://biostatgv.sentiweb.fr/> ou encore pvalue.io www.pvalue.io qui proposent des solutions en ligne pour réaliser facilement des analyses statistiques. Ces derniers sont cependant très focalisés sur la réalisation de tests d'hypothèse qui ne constituent qu'une partie de l'analyse des données. Dans *GMRC Shiny Stats*, la mise en page sous forme d'onglets définit une progression logique qui permet à l'utilisateur de réaliser et de comprendre l'intérêt des analyses descriptives dans le processus d'analyse et de (re)-situer le test d'hypothèse comme une composante parmi d'autres permettant de mener à des conclusions sur les questions à résoudre. Ce repositionnement imposé par la forme est également facilité par la disparition de l'angoisse du choix du « bon » test puisque l'application guide l'utilisateur vers l'utilisation du test le mieux adapté en fonction des données. Ainsi, en comparaison à d'autres solutions logicielles disponibles qui ne permettent pas de sélectionner le test statistique à utiliser dans une situation donnée et où c'est à l'utilisateur de réaliser ce choix, *GMRC Shiny Stats* apporte un vrai bénéfice pour les utilisateurs qui ne se sentent pas à l'aise avec cette étape de l'analyse.

Un autre avantage de *GMRC Shiny Stats* porte sur la confidentialité des données. Les données personnelles de santé sont des données sensibles. Même pseudonymisées, ces données restent relativement sensibles car il est difficile d'assurer que les étapes d'anonymisation ou de pseudonymisation soient totalement irréversibles (21–23). Contrairement à des serveurs web qui ne bénéficient pas nécessairement de l'agrément hébergeur de santé, la sécurité des données est totalement assurée dans *GMRC Shiny Stats* car les données restent uniquement à leur emplacement initial et aucun enregistrement des données n'est réalisé. Le prix à payer est qu'il n'est pas possible d'enregistrer sa session *GMRC Shiny Stats* - il est donc nécessaire de recharger la base de données à chaque fois que l'on reprend une session de travail.

Une des limites des analyses faites à l'aide de *GMRC Shiny Stats* et qui résulte de la nature entièrement cliquable de l'application est relative aux bonnes pratiques en terme de reproductibilité des analyses. La recherche reproductible repose sur l'idée que les analyses de données, et plus généralement les allégations scientifiques, sont publiées avec leurs données et leur code logiciel afin que d'autres puissent vérifier les résultats et les exploiter (24). Bien que le code du logiciel soit ouvert, l'absence d'instructions sous forme de lignes de commandes limite la reproductibilité des analyses. L'application ne permet

donc pas de satisfaire aux bonnes pratiques en termes de reproductibilité puisque les actions des utilisateurs ne sont pas tracées. Le choix de proposer une application didactique s'est faite au détriment de la reproductibilité des analyses.

Nous avons également fait le choix de ne pas aborder les analyses de dimensionnalités supérieures et l'outil ne permet donc pas de réaliser d'analyses multivariées. Il s'agit d'un choix délibéré qui repose sur l'idée que l'utilisation de modèles multivariés devient très rapidement complexe et dépasse souvent le niveau correspondant à la formation de base des étudiants. L'utilisateur novice qui serait tenté par ce genre d'approches gagne à bénéficier de l'expertise d'un statisticien confirmé. La proposition de réaliser des analyses statistiques uniquement selon des approches fréquentistes est également discutable puisque les enseignements qui abordent les statistiques bayésiennes sont de plus en plus nombreux et l'inférence bayésienne est possiblement conceptuellement plus simple à comprendre pour les étudiants (25, 26). Une version permettant de réaliser des analyses de données médicales à l'aide de méthodes bayésiennes est actuellement en cours de développement.

En conclusion, *GMRC Shiny Stats* permet aux étudiants de participer de manière active et de manière autonome à la réalisation d'analyses de données médicales, notamment dans le cadre des travaux universitaires. Notre expérience dans l'accompagnement de ces étudiants nous montre que c'est pour eux un moyen facile de comprendre un certain nombre de concepts statistiques. Se confronter aux problématiques et aux éventuelles limites de l'analyse statistique des données dans les processus de recherche - même à l'aide d'un outil simplifiant très largement la démarche - permet l'acquisition de compétences visant à mieux comprendre, communiquer et vulgariser les concepts biostatistiques utilisés dans le domaine médical.

REMERCIEMENTS

Nous souhaitons remercier tous les utilisateurs de *GMRC Shiny Stats*, en particulier les adopteurs précoces qui nous ont fait remonter les inévitables bugs ou coquilles. Nous remercions également les relecteurs de l'article lors de la soumission de cet article pour la qualité de leur travail de relecture et la pertinence de leurs remarques et commentaires.

Bibliographie

1. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*, 268(17):2420–5, nov 1992. ISSN 0098-7484. doi: 10.1001/jama.1992.03490170092032.
2. Gerd Gigerenzer, Wolfgang Gaissmaier, Elke Kurz-Milcke, Lisa M. Schwartz, and Steven Woloshin. Helping Doctors and Patients Make Sense of Health Statistics. *Psychological Science in the Public Interest*, 8(2):53–96, nov 2007. ISSN 1529-1006. doi: 10.1111/j.1539-6053.2008.00033.x.
3. Douglas G. Altman and J. Martin Bland. Improving Doctors' Understanding of Statistics. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 154(2):223, 1991. ISSN 09641998. doi: 10.2307/2983040.
4. Thierry Pelaccia. Comment (mieux) superviser les étudiants en sciences de la santé dans leurs stages et dans leurs activités de recherche ? *Pédagogie Médicale*, 19(1):51–52, feb 2018. ISSN 1625-6484. doi: 10.1051/pmed/2019006.
5. Winston Chang, Joe Cheng, J J Allaire, Yihui Xie, and Jonathan McPherson. *shiny: Web Application Framework for R*, 2019.
6. David A Grimes and Kenneth F Schulz. Descriptive studies: what they can and cannot do. *The Lancet*, 359(9301):145–149, jan 2002. ISSN 01406736. doi: 10.1016/S0140-6736(02)07373-7.
7. Douglas G Altman and J Martin Bland. Missing data. *BMJ (Clinical research ed.)*, 334(7590):424, feb 2007. ISSN 1756-1833. doi: 10.1136/bmj.38977.682025.2C.
8. Allison Shorten and Brett Shorten. Which statistical tests should I use? *Evidence-based nursing*, 18(1):2–3, jan 2015. ISSN 1468-9618. doi: 10.1136/eb-2014-102003.
9. Evie McCrum-Gardner. Which is the correct statistical test to use? *British Journal of Oral*

- and *Maxillofacial Surgery*, 46(1):38–41, jan 2008. ISSN 02664356. doi: 10.1016/j.bjoms.2007.09.002.
10. Rink Hoekstra, Henk Kiers, and Addie Johnson. Are Assumptions of Well-Known Statistical Techniques Checked, and Why (Not)? *Frontiers in Psychology*, 3, 2012. ISSN 1664-1078. doi: 10.3389/fpsyg.2012.00137.
 11. Izet Masic, Slobodan M Jankovic, and Edin Begic. PhD Students and the Most Frequent Mistakes During Data Interpretation by Statistical Analysis Software. *Studies in health technology and informatics*, 262:105–109, jul 2019. ISSN 1879-8365. doi: 10.3233/SHTI190028.
 12. Terry M Therneau and Patricia M Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer, New York, 2000. ISBN 0-387-98784-3.
 13. J Martin Bland and Douglas G Altman. The logrank test. *BMJ (Clinical research ed.)*, 328(7447):1073, may 2004. ISSN 1756-1833. doi: 10.1136/bmj.328.7447.1073.
 14. Xavier Robin, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, and Markus Müller. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12:77, 2011.
 15. Jacob Cohen. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46, apr 1960. ISSN 0013-1644. doi: 10.1177/001316446002000104.
 16. Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382, 1971. ISSN 0033-2909. doi: 10.1037/h0031619.
 17. J R Landis and G G Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–74, mar 1977. ISSN 0006-341X.
 18. Mousumi Banerjee, Michelle Capozzoli, Laura McSweeney, and Debajyoti Sinha. Beyond kappa: A review of interrater agreement measures. *Canadian Journal of Statistics*, 27(1): 3–23, mar 1999. ISSN 03195724. doi: 10.2307/3315487.
 19. Rui Wang, Stephen W Lagakos, James H Ware, David J Hunter, and Jeffrey M Drazen. Statistics in medicine—reporting of subgroup analyses in clinical trials. *The New England journal of medicine*, 357(21):2189–94, nov 2007. ISSN 1533-4406. doi: 10.1056/NEJMs077003.
 20. Stephen W. Lagakos. The Challenge of Subgroup Analyses — Reporting without Distorting. *New England Journal of Medicine*, 354(16):1667–1669, apr 2006. ISSN 0028-4793. doi: 10.1056/NEJMp068070.
 21. Rik Crutzen, Gjalt-Jorn Ygram Peters, and Christopher Mondschein. Why and how we should care about the General Data Protection Regulation. *Psychology health*, pages 1–11, may 2019. ISSN 1476-8321. doi: 10.1080/08870446.2019.1606222.
 22. April Moreno Arellano, Wenrui Dai, Shuang Wang, Xiaoqian Jiang, and Lucila Ohno-Machado. Privacy Policy and Technology in Biomedical Data Science. *Annual review of biomedical data science*, 1:115–129, jul 2018. ISSN 2574-3414. doi: 10.1146/annurev-biodatasci-080917-013416.
 23. Mehmet Kayaalp. Patient Privacy in the Era of Big Data. *Balkan Medical Journal*, 35(1): 8–17, jan 2018. ISSN 21463123. doi: 10.4274/balkanmedj.2017.0966.
 24. Roger D. Peng. Reproducible research and Biostatistics. *Biostatistics*, 10(3):405–408, jul 2009. ISSN 1468-4357. doi: 10.1093/biostatistics/kxp014.
 25. Stephanie Kurzenhäuser and Ulrich Hoffrage. Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. *Medical teacher*, 24(5):516–21, sep 2002. ISSN 0142-159X. doi: 10.1080/0142159021000012540.
 26. N Meyer, S Vinzio, and B Goichot. [Bayesian statistic: an approach fitted to clinic]. *La Revue de médecine interne*, 30(3):242–9, mar 2009. ISSN 0248-8663. doi: 10.1016/j.revmed.2008.07.004.