*B*ulletin de la *D*ialyse à *D*omicile

Development of an application for accessing and analyzing RDPLF data

(Développement d'une application d'accès et d'analyses des données du RDPLF)

https://shiny.rdplf.org/rdplf-hdd/

Martin Masson[1], Benoît Thieurmel[2]

[1]Data Scientist R&D, Datastorm, [2]Directeur des opérations, Datastorm

Résumé

La collecte d'informations dans la base de données d'un registre médical trouve son premier intérêt dans la possibilité pour le médecin et l'équipe soignante d'analyser ses résultats et se comparer avec d'autres équipes aux fins de partage d'expérience et connaissances. Le RDPLF collige depuis 1986 les données de 45 000 patients insuffisants rénaux traités à domicile dans les pays francophones. Un partenariat a été créé entre le RDPLF et Datastorm (https://www.datastorm.fr), la filiale d'expertise et de conseil du Groupe ENSAE-ENSAI (Ecoles nationales d'économie et statistique) afin de développer une application qui permette à tout professionnel de santé, au moyen d'un interface WEB convivial, de réaliser des requêtes statistiques simples sur la base de données du RDPLF. Ainsi, tout médecin ou membre d'équipe soignante peut évaluer, sans compétences statistiques spéciales, les résultats par région et pays francophones. Un accès spécial permet également à tout centre de confronter ses propres résultats à ceux d'une région de référence. Les graphiques générés sont utilisables pour des présentations lors de réunions d'équipe ou pour des travaux. Le logiciel support de l'application est basé sur le logiciel R (https://www.r-project.org) et son interface de visualisation SHINY (https://shiny.rstudio.com). Nous rapportons les modalités de développement de cette application, ses fonctionnalités (basées sur des critères présélectionnés : taux d'incidence, prévalence, survie, taux d'infections, répartition des traitements, aspects infirmiers). Cet article décrit comment effectuer ces études de manière abordable à la fois aux infirmières et médecins. Son interface bilingue l'ouvre également à la communauté anglophone et facilite ainsi la communication internationale.

Summary

The collection of information in the database of a medical registry finds its first interest in the possibility for a doctor and a care team to analyze their results and to compare themselves with other teams for the purpose of sharing experience and knowledge. Since 1986, the French Language Peritoneal Dialysis Registry (RDPLF) has collected data from 45,000 patients with renal failure treated at home in French-speaking countries. A partnership has been created between the RDPLF and Datastorm (https://www.datastorm.fr), the expertise and consultancy subsidiary of the ENSAE-ENSAI Group (National Schools of Economics and Statistics) in order to develop an application that allows to carry out simple statistical queries on the RDPLF database, by means of a user-friendly WEB interface. Thus, any doctor or member of the healthcare team can evaluate, without any special statistical skills, results by region and by French-speaking country. Special access also allows any center to compare its own results with those of a reference region. The generated graphics can be used for presentations during team meetings or for work. The application is based on the R software (https://www.r-project.org) and its SHINY visualization interface (https://shiny.rstudio.com). We report on development modalities and its functionalities (based on preselected criteria: incidence rate, prevalence, survival, infection rate, distribution of treatments, nursing aspects). This article describes how both nurses and doctors can easily realize studies with the application. Its bilingual interface also opens it up to English-speaking communities and thus facilitates international communication.

Correspondance : Benoit Thieurmel, 24 rue Barbès
92120 MONTROUGE - FRANCE

journal officiel du Registre de Dialyse Péritonéale de Langue Française   **RDPLF**   www.rdplf.org

INTRODUCTION

Important note:
The figures and graphics presented in this article are examples with fictitious selection criteria. They should not be copied or cited as is.

Since its creation in 1986, the RDPLF has been collecting peritoneal dialysis data from nearly 45,000 peritoneal dialysis patients from French-speaking countries. The database is structured in several modules: a compulsory main module including monitoring of comorbidities, assistance, infections, survival, transplantation, and optional modules specialized in more specific aspects: nursing, catheter follow up, anemia, nutrition, heart failure. [1]

The RDPLF is Registered in the CNIL (National Commission of Information and Liberty) under the numero : 542668

Since 2012, patients treated at home with hemodialysis have also been progressively added to the database. While the database includes 98% of peritoneal dialysis patients, it only includes 40% of home hemodialysis patients. However, with the development of home dialysis machines, it is very important to be able to monitor and analyze these new patients' data.
Thus, the RDPLF continually maintains and enriches its database and allows medical and nursing staff to access and analyze it, in several ways :

- *By writing annual summaries on its website [2]* or in the Bulletin de la Dialyse à Domicile [3]
- *By providing web applications* (developed with the R language and the shiny module) describing:
	+Autonomy: distribution of levels of autonomy and types of assistance for prevalent patients
	+ Profile: center activity and recruitment profile
	+PD discharge: Discharge mode for patients who have been treated with peritoneal dialysis
	+Survival: Technical survival and patient survival curves
- *By an interface allowing to freely query* the data and to carry out various statistical analyses by sub-population via the development of php routines (http://www.rdplf.org/rdplf_db/index.php).

The RDPLF approached Datastorm at the end of 2019 for support in sharing and improving on the various existing tools.

An expertise and consulting subsidiary of the ENSAE-ENSAI Group, Datastorm enables companies and public organizations to enrich their own Research and Development capacity by mobilizing very rich expertise in Data Science, Artificial Intelligence and Economics from the laboratories of the ENSAE-ENSAI Group and more broadly from the Polytechnic Institute of Paris.

The roadmap was as follows:
- Unify the four R-Shiny web applications put online by the RDPLF
- Include existing statistical querying functionalities
- Add new analysis modules, such as the «nurse» module
- Make the application available in French and English
- Improve the overall ergonomics

journal officiel du Registre de Dialyse Péritonéale de Langue Française   **RDPLF**   www.rdplf.org

TECHNICAL ENVIRONMENT

The RDPLF already had applications developed with the R language (https://www.r-project.org/about.html) and the shiny package (https://shiny.rstudio.com/), as well as an architecture to host them and make them accessible. The new web application therefore uses the same technologies.

All the data of the different modules is stored in a MySQL database.
To add interactivity to charts and table visualizations, the following two R packages were used:

- rAmCharts: a package which enables all the visualizations offered by the amCharts Javascript library (https://www.amcharts.com/) to be created in the R environment.
- DT: interface with the Datatables.js library for tables (https://rstudio.github.io/DT/).

In addition, we used the shinymanager package (https://github.com/datastorm-open/shinymanager) to enable secure authentication and rights management in the application.

Note that Datastorm has chosen to contribute to the open source ecosystem, notably by developing several R and python packages to advance data science. rAmCharts and shinymanager are among these packages available on the company's GitHub page (https://github.com/datastorm-open). The development of the RDPLF application has also made it possible to identify new needs and integrate them into these packages.
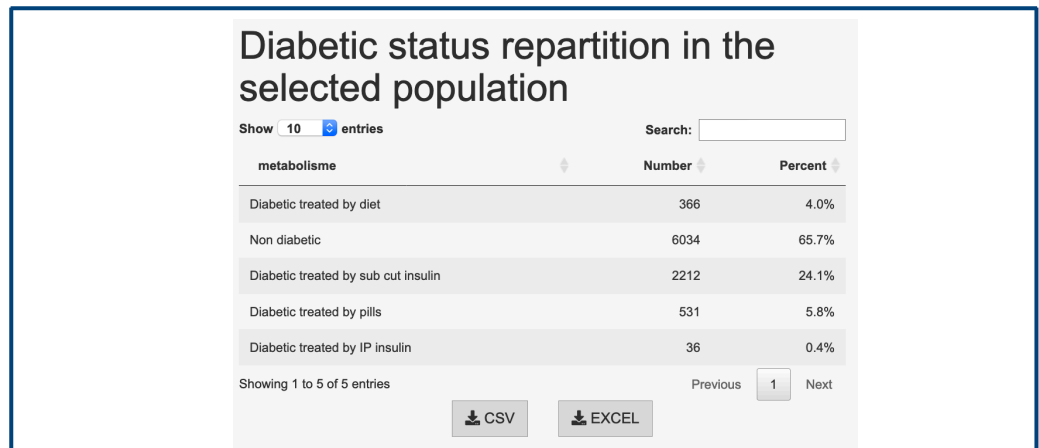
On all the results presented in the application, it is therefore now possible:

1) To generate an interactive display of values by passing the mouse over the graph (Fig. 1)
2) To zoom in on a sub-period and display only the desired information by clicking on the legend
3) Export the graphic (.png, .jpg, .pdf) using the button at the top right
(exemples in figure below)
4)View and download tables (.csv, .xlsx) by clicking on the «View table» button. Smaller arrows in the header allow you to sort the table according to the chosen column.(Fig.2).

REFACTORING THE EXISTANT
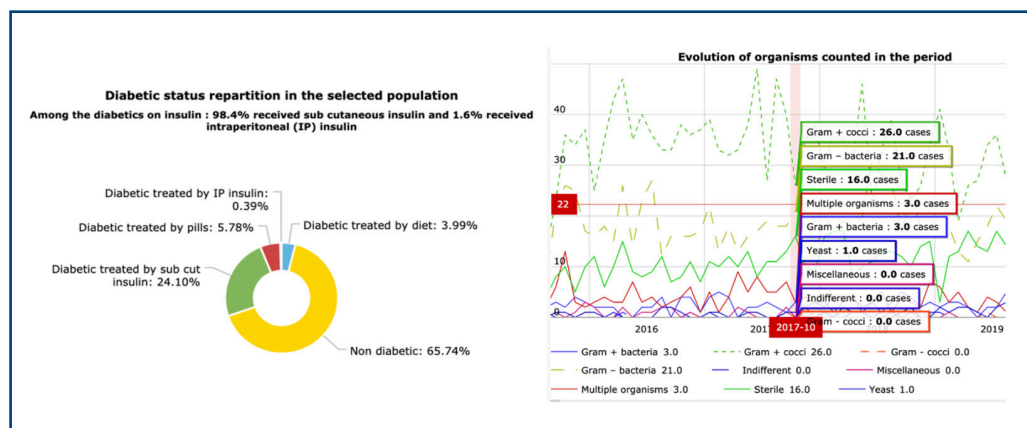
⬆Fig. 1: screenshots on the left and on the right showing the legends that appear automatically when the mouse passes over the graph: allows you to view the value of each point on the graph.

Diabetic status repartition in the selected population

metabolisme	Number	Percent
Diabetic treated by diet	366	4.0%
Non diabetic	6034	65.7%
Diabetic treated by sub cut insulin	2212	24.1%
Diabetic treated by pills	531	5.8%
Diabetic treated by IP insulin	36	0.4%

Show 10 entries   Search:

Showing 1 to 5 of 5 entries   Previous 1 Next

⬆ CSV   ⬆ EXCEL

⬆*Fig. 2: window allowing the downloading of graph data in the form of a CSV or EXCEl file to allow the user to reuse them as needed.*

The first step made it possible to integrate all the existing functionalities previously developed by RDPLF, namely all the php routines allowing to perform various free analyses by sub-population, as well as the four web applications (Autonomy, Profile, DP output and Survival).

The data from the original database is automatically exported every night, after complete anonymization, to a MySQL database on a different server unrelated to the original server, so that it is not possible to access the original personal information. All the queries are carried out on this remote and independent MySQL database.

I. Patients selection

By connecting to the application, one must first select the scope of patients to be analyzed, according to the following criteria (fig.3):

- Location: countries / regions or directly a center according to user profiles
- Inclusion period
- Patients' type: prevalent or incidents
- Treatments before peritoneal dialysis
- Age group as of the first peritoneal dialysis
- Gender
- Charlson's comorbidity index
- Diabetic status
- Autonomy type(s)
- Peritoneal dialysis type

Data is collected and the results are generated once the various criteria are validated by clicking on the button «Click to validate the selection criteria» present above the filters.

In addition, there is a quick descriptive view of the whole RDPLF database to date, with the total number of patients and centers since its existence(Fig.4) :

It is possible to change the scope of the study at will by returning to the «Selection criteria» tab at the top of the banner on the left. L A button displays the details of the current sub-population, linked to the results displayed in the rest of the application.. (Fig.5)

⬆Fig. 3: Window of the selection criteria that will be applied to the various calculations
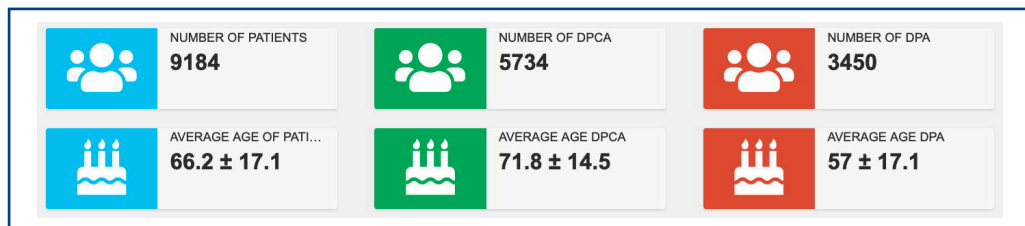

⬆Fig. 4: On the home page, independent of the selection criteria, a reminder of the total number of patients and centers included from the start (all centers, all ages, sex and comobidities combined)

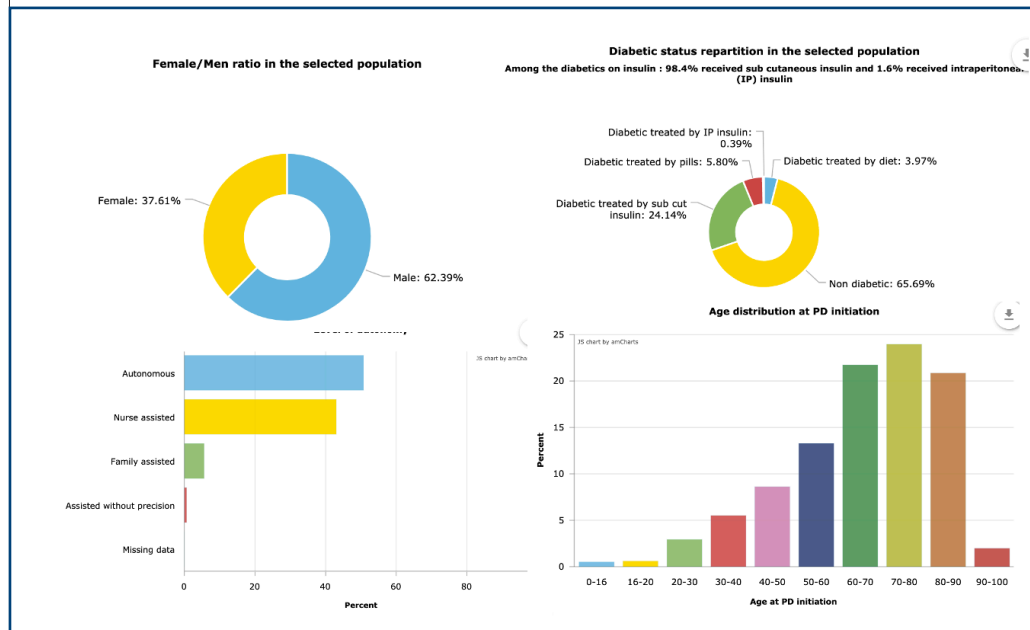⬆Fig. 5: reminder of the criteria selected for the rest of the analyzes

II. Description of the selected population

- A first banner (fig.6) containing the following key figures:
    + Number and age (mean = standard deviation) of all patients
    + The APD / CAPD detail
- A series of four graphics detailing (fig.7):
    + Female to male ratio
    + The distribution of diabetic status
    + Levels of autonomy
    + The age pyramid at dialysis start

III. Patients movements

journal officiel du Registre de Dialyse Péritonéale de Langue Française **RDPLF** www.rdplf.org

⬆*Fig. 6: General description of the selected population*

⬆*Fig. 7: Gender distribution, autonomy, diabetic status and ages*

The second tab focuses on patient movements. One finds in a first banner (fig.8) the following information:

- Number of new patients
- Patients at the start and end of the period
- Transplanted / deceased patients
- The number of transfers in hemodialysis

Completed with a new series of charts representing the distribution (fig.9):
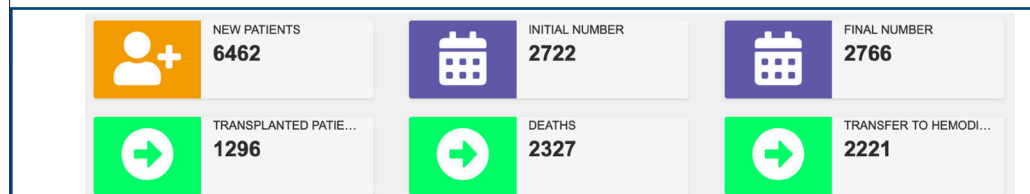
⬆*Fig. 8: Summary of patient movements*

- dropout causes
- Causes of transfers in hemodialysis
- Causes of death
- waiting times before transplantation

Finally, note that this menu is only available in the case of a selection of prevalent patients.

IV. Duration on dialysis (figures 10 a,b,c)
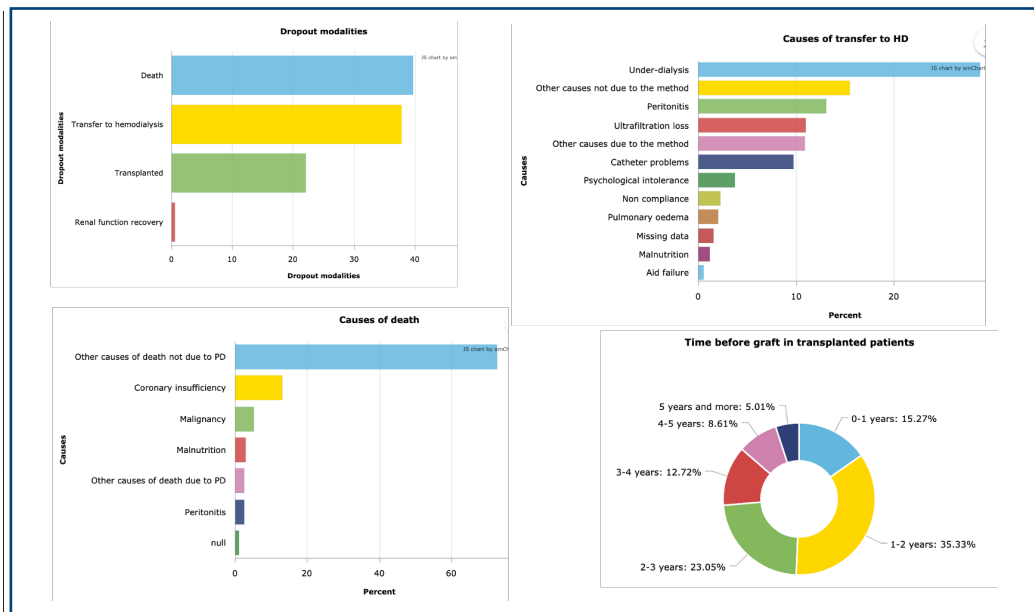
Fig. 9: peritoneal dialysis dropout modalities

In the third tab, we find the study of dialysis durations, with three series of curves:
- Patient and technical survival of incident patients
- Probabilities of transfers to hemodialysis and transplantation of incident patients :•   •
- Cumulative incidence of peritoneal dialysis events

Unlike the last graph showing the cumulative incidence, the first two do not take competitive risks into account [5,6]. They overestimate the events of interest and thus offer an estimate of the theoretical impact, that is, an estimate in a setting with an absence of concurrent events. Models used are Cox models, with competitive event censorship. Overall technical survival is survival based on a composite endpoint defined by ANZDATA [4] (death or HD transfer). Strict technical survival only takes transfers into account.

The method used to calculate the confidence intervals is the 'plain scale' method. The reader will find more information in the following document: http://iml.univ-mrs.fr/~reboul/R-survie.pdf
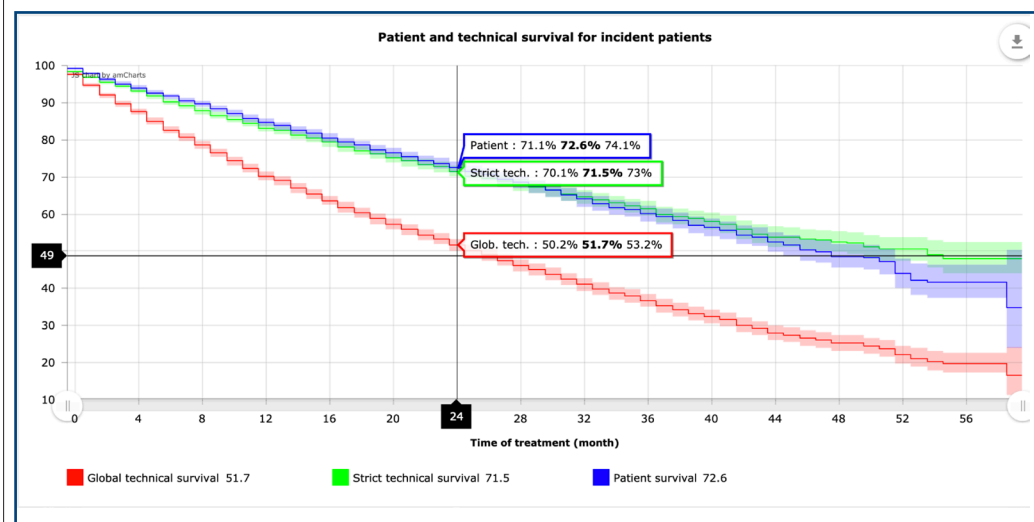
Fig. 10a: patient and technique survival
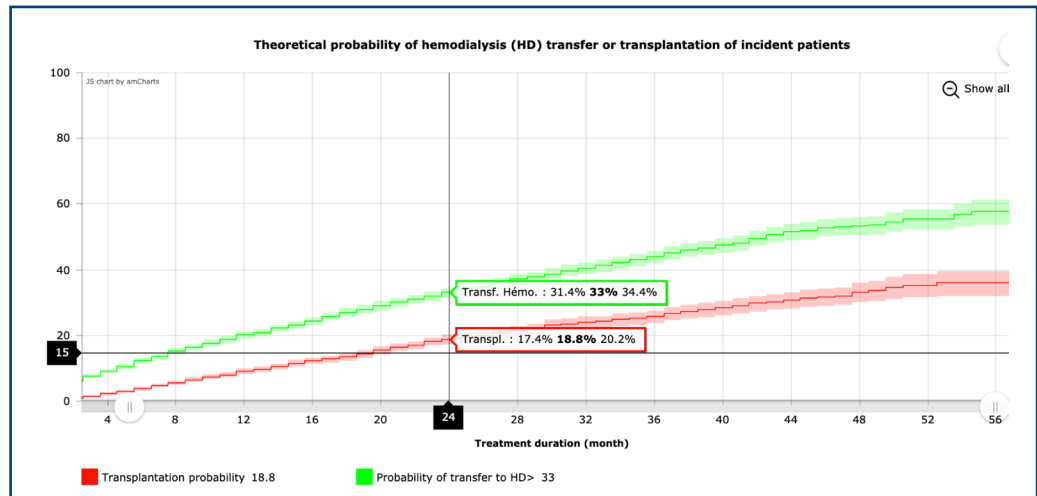
Fig. 10b: acturarial probability of transfer to hemodialysis and transplantation
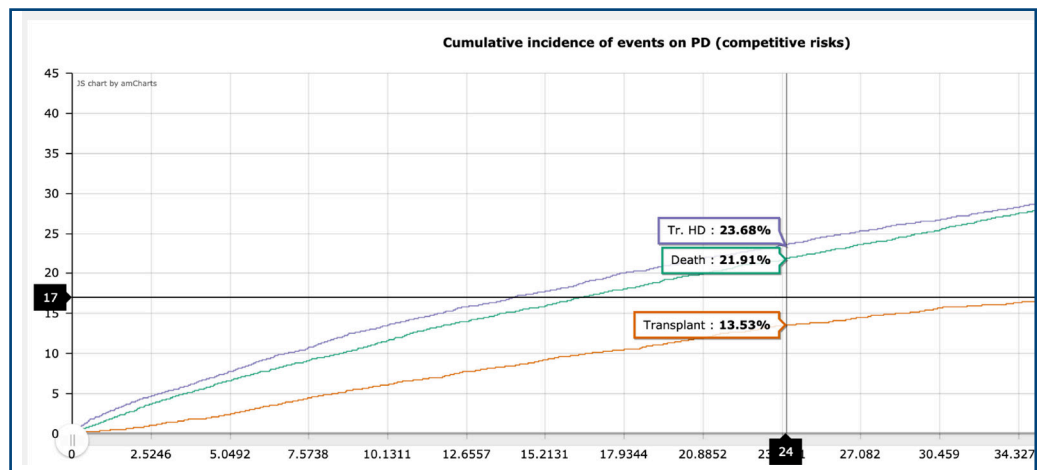
Fig.10c: cumulative incidence of different PD dropout modalities taking into account competitive risks

V. Study of peritonitis

The last menu of the main module focuses on the study of peritonitis with, first, a table detailing episodes' frequency (Figure 12):

• The number and percentage observed over the period
• Rates in months x patients / number and years x patients
• All detailed by type of peritonitis

We then find information on the ecology of organisms involved in peritoneal infection, in two levels of detail:

By families of organisms: distribution of families over the period (fig. 13) and monthly rate of

Peritonitis type	Count during the period	Percent during the period	Rate (month x patients / count)	Rate (year x patients)
Everything	4070	1	38.18	0.31
Everything except relapses (ISPD)	3898	0.9577	39.86	0.3
1st Peritonitis	1699	0.4174	91.46	0.13
New peritonitis	948	0.2329	163.91	0.07
Recurrence	156	0.0383	996.04	0.01
Relapse	239	0.0587	650.14	0.02
Repeat	172	0.0423	903.39	0.01

Fig. 12: Frequencies of peritoneal infections, in detail, frequency of the 1st episode, total frequency, frequency of recidives, frequency of recurrences (according to ISPD definition)

each type of infection (fig 14)
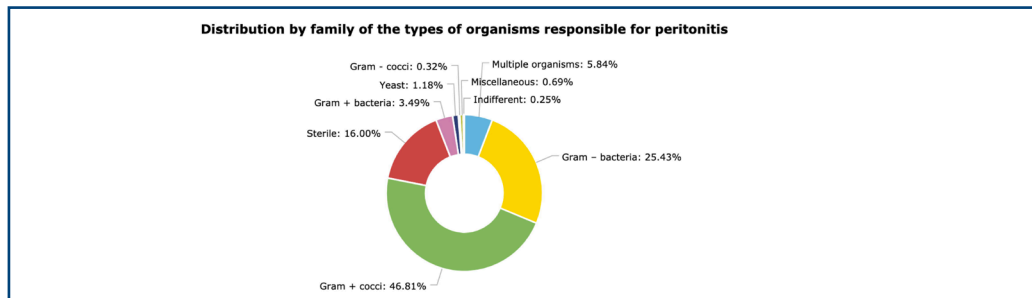
⬆Fig. 13: Distribution of the main types of otganisms.
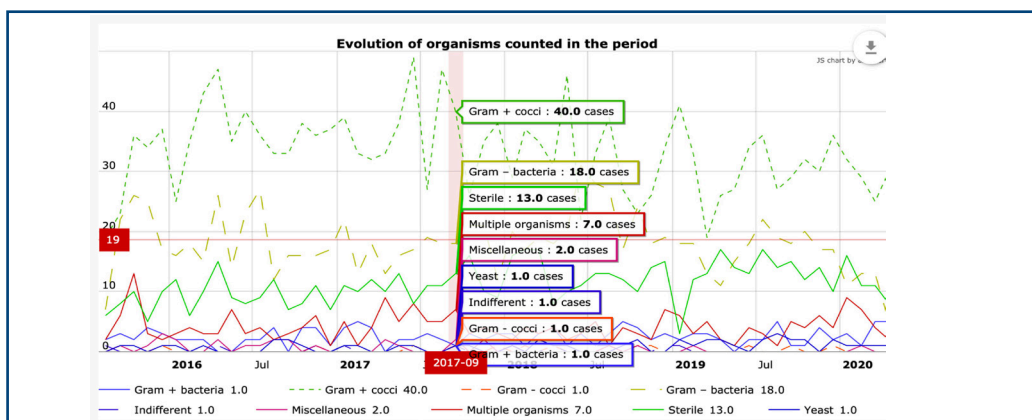
⬆Fig. 14: temporal evolution of ecology. Hovering the mouse over the curves allows the numerical display of variables in number or percentage. The time display thus makes it possible to detect an abnormal variation in the ecology.

At the level of organisms: we can focus on one or more families of organisms and obtain their count and evolution over time.

All views of the bacterial ecology are available on the desired type (s) of peritonitis. To do this, one must simply select the corresponding lines in the first table (fig.15) showing the frequency of episodes, fig.16 is then automatcally updated.

Family	Name	Number	Percent
Gram + cocci	Enterococcus faecalis	154	3.8%
Gram + cocci	S. aureus (non précisé)	203	5.0%
Gram + cocci	S. aureus méti-S	215	5.3%
Gram + cocci	S. epidermidis (non précisé)	244	6.0%
Gram + cocci	S. epidermidis méti-S	131	3.2%
Gram + cocci	Staphylococcus haemolyticus	109	2.7%
Gram – bacteria	Escherichia Coli	281	6.9%
Multiple organisms	Germes multiples	240	5.9%

⬆Fig. 15: table of the different organisms recorded in the selected period. By clicking on one or more lines, the temporal evolution of figure 16 is updated automatically.
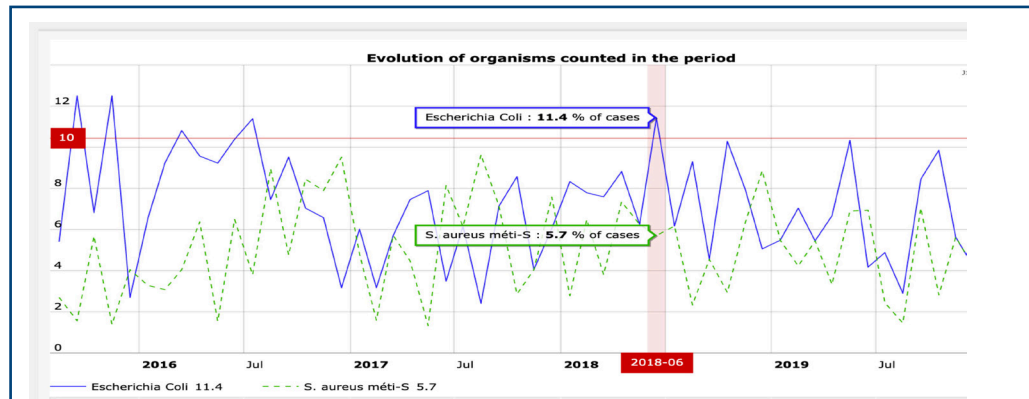
⬆ *Fig. 16: Evolution of the number or percentage of organisms selected in Figure 15. In the present case, we see an increase in the percentage of Eschericha coli in June 2018 (fictiv data for the example, do not use this figure)*

NURSING MODULE

In addition, an interface allowing the exploration and analysis of the nursing module was developed. This possibility was not available before. This module is linked directly to the population selected upstream, but only on the location filters (Country, regions, centers) and on the inclusion period. We focus here only on incident patients.

I. Contextual informations

The first page provides contextual information with the number and percentage of all centers (resp. Patients) present in the scope of the module. (fig. 17) .
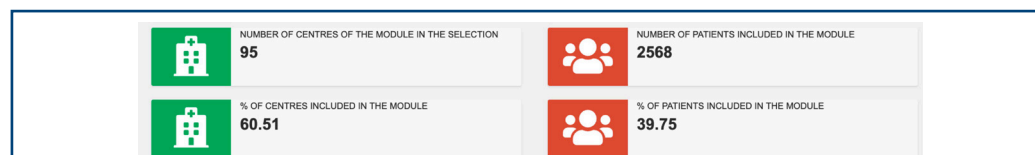
⬆ *Fig. 17: This banner reminds you of the completeness of the data in relation to all the patients*

Moreover, for illustrative purposes, we find the total number of nursing modules selected per year. This information is important as data on nursing practices are less exhaustive in the original data base.(fig. 18).
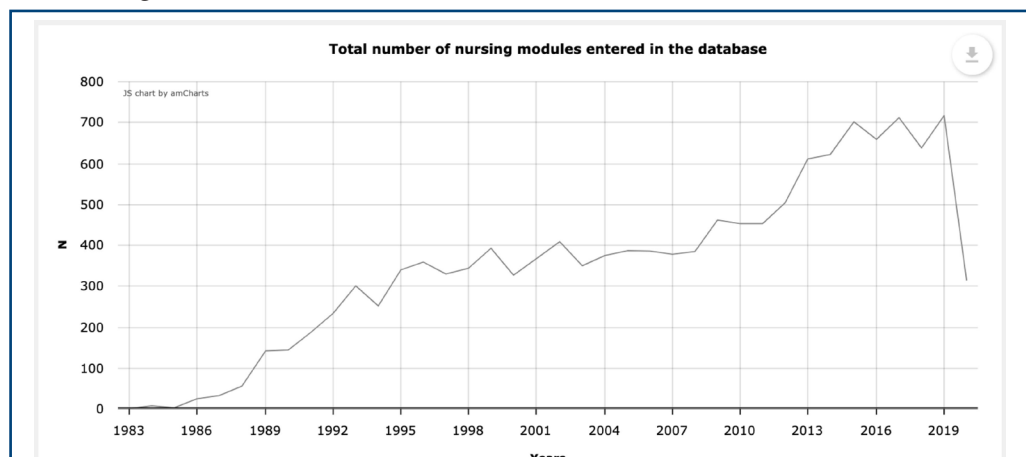

⬆ *Fig. 18: Total number of nursing modules registered in the database, year by year. These are only patients from the centers that were initially selected and only concern incident patients in each period.*

journal officiel du Registre de Dialyse Péritonéale de Langue Française RDPLF www.rdplf.org

II. Care training

We then focus on healthcare training, with:
- • The distribution of the waiting times between catheter placement and the start of training
- • The percentage of patients trained before and after catheter placement
- • The distribution of the duration of training
- • The people following training
- • Details of the training materials used
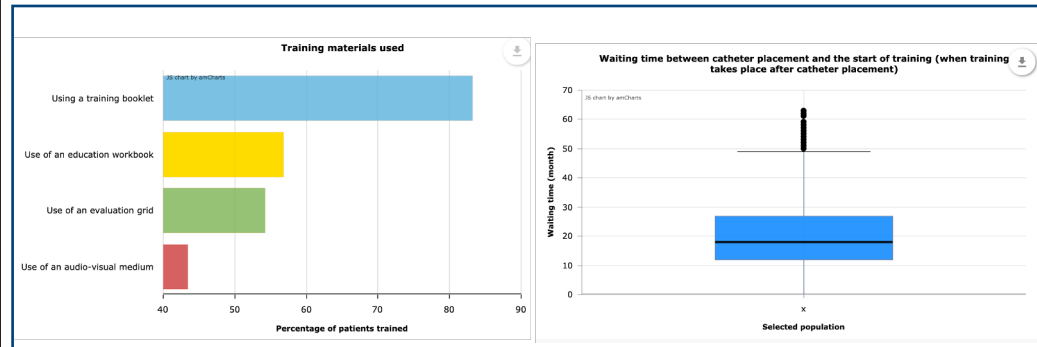
⬆ *Fig. 19: Examples of graphs of the nursing module concerning training support and time between catheter implantation and initiation of training.*

III. Study of disabilities

The focus in the third tab is on patients with disabilities, with different distributions among selected patients (fg.20):
- • Level of intellectual alteration
- • Visual and hearing impairments
- • Method of learning the injection technique
- • Percentage of patients with a disability at the time of treatment
- • How to choose the dialysis method
- • Patient positions regarding the injection technique

IV. Autonomous patients
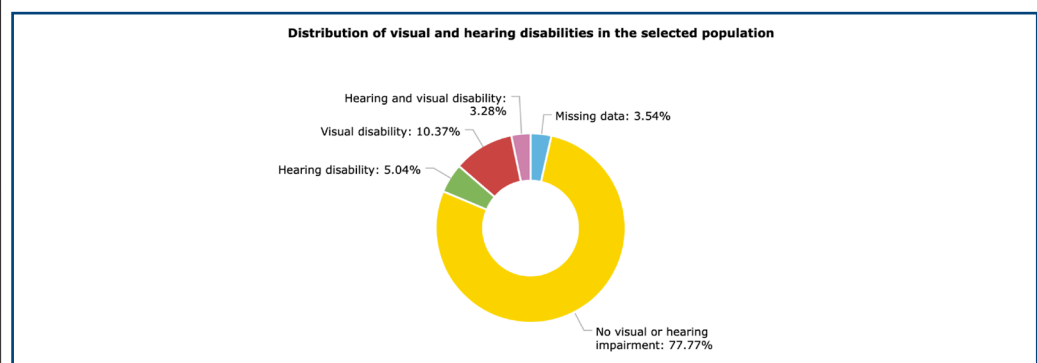
⬆ *Fig. 20: Frequency of the different disabilities in the selected population*

Finally, we find information on autonomous patients, namely their ability to adapt their bags and measure their blood pressure (Fig.21).
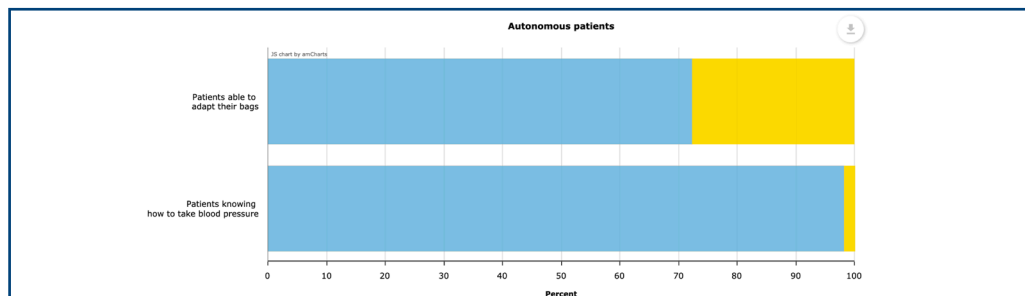
⬆*Fig. 21: distribution of patients who know how to change their bags and take their blood pressure*

MUTILINGUAL DEVELOPMENT

The applicaton is available both in French and English (fig. 22).
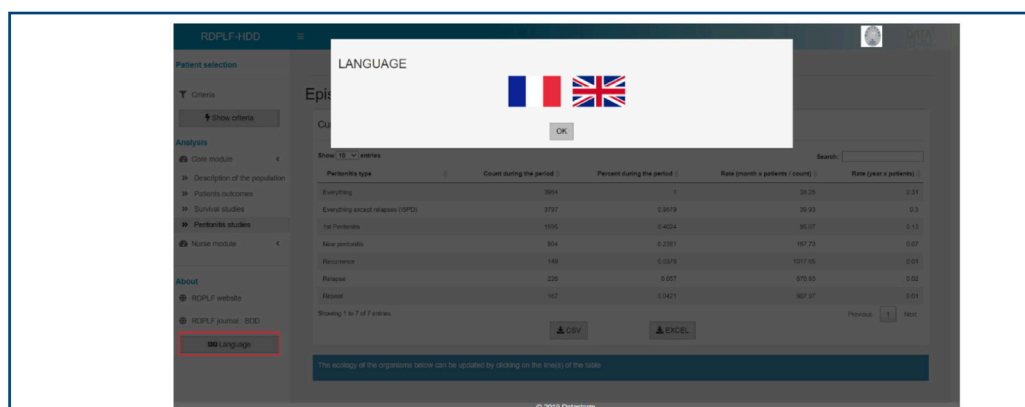
⬆*Fig. 22: It is possible to change the language at any time by using the "Language" button at the bottom of the left menu.*

SECURING THE APPLICATIO AND MANAGING USERS

The application is freely available to all physicians and nurses interested in home dialysis. To date, the application is not freely accessible to general public as its use needs at least basic knowledge in epidemiology and statistics: visitors who would not have the background could misinterpret some of the results. But anyone can ask access to the RDPLF Staff by indicating his or her interest. The RDPLF decides then the access conditions. The access may become totally public in the future, depending on the returns on the current version.

Access User authentication and rights management are done securely, with the shinymanager package (https://github.com/datastorm-open/shinymanager) developed by Datastorm.

Indeed, user data is stored in a SQLite database encrypted with the AES (Advanced Encryption Standard) symmetric encryption algorithm and the openssl library (https://www.openssl.org/). In addition, passwords are hashed.

Access to the application is via https (HyperText Transfer Protocal Secure), all data sent by the user to the server is thus secure.

To respect RGPD recommendations: no analysis nor user profiling are performed on the entered information. It is reserved for internal use and not communicated to anyone else. At any moment, regsitered visitors can ask access to this information and can ask to be deleted.

Users can also be attached to one or more centers or associations. If this is the case, they can then choose to study only the patients of their center (s) / association (s) or to define a geographical perimeter.
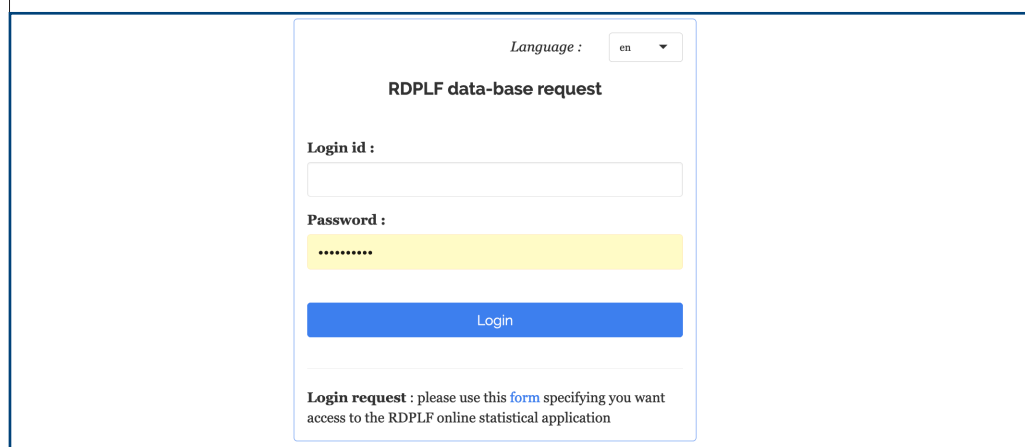
ACCESS TO THE APPLICATION

The application is accessible at the following address (Fig. 23):
https://shiny.rdplf.org/rdplf-hdd/

The link is also available from the RDPLF site (https://www.rdplf.org) in the statistics / personal requests menus)

Currently, it is not possible to register directly from the app. One must first make a registration request using a contact form specifying «RDPLF statistical application access»: https://www.rdplf.org/contact.html

⬆*Fig. 23: application login screen and registration link*

CONCLUSION

Developed by the Datastorm teams on behalf of the RDPLF, the RDPLF-HDD application is a new work and reflection tool made available to professionals, nephrologists and nurses specializing in home dialysis. It allows them to analyze data from the RDPLF database and export multi-criteria results in a simple, dynamic and interactive way. Depending on its use and the needs and demands identified, it will be possible to integrate other analysis modules in the service of medical knowledge and expertise. To date only access to peritoneal dialysis has been developed: future improvements could include home hemodialysis.

journal officiel du Registre de Dialyse Péritonéale de Langue Française   **RDPLF**   www.rdplf.org

which made this project possible.

REFERENCES

1 - Verger C, Ryckelynck JP, Duman M, et al. French peritoneal dialysis registry (RDPLF): outline and main results [published correction appears in Kidney Int Suppl. 2007 Jan;71(1):87. Fabre, E [added]]. Kidney Int Suppl. 2006;(103):S12-S20. doi:10.1038/sj.ki.5001911

2-Statistiques 2017 :   https://rdplf.org/resultatsrdplf/epidemiologie-dialyse-peritoneale.html (consulté 27/7/2020)

3 -Verger C, Fabre E, Veniez G, et Padernoz MC. 2019. « Données Statistiques synthétiques Du Registre De Dialyse Péritonéale De Langue Française et Hémodialyse à Domicile (RDPLF) En 2018 ». Bulletin De La Dialyse à Domicile 2 (1), 1-10. https://doi.org/10.25796/bdd. v2i1.19093. (consulté 27/7/2020)

4 - Lan PG, Clayton PA, Johnson DW, McDon-ald SP, Borlace M, Badve SV, et al. Duration of hemodialysis following peritoneal dialysis cessation in Australia and New Zealand: proposal for a standardized definition of tech-nique failure. Perit Dial Int. 2016; 36(6): 623–630

5. Gray R-J. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk, The Annals of Statistics, Vol. 16, No. 3 (Sep., 1988), pp. 1141-1154.

6. Sandrine Katsahian,Hugo Varet. Incidence cumulée et risques compétitifs : ça veut dire quoi ?Horizons Hémato 2012; 2(2) http://horizonshemato.com/wp-content/uploads/2017/10/ Horizons-Hemato-Juin-2012-108-109-Katsahian-Varet.pdf (consulted 27/7/2020)