Interactions entre les évènements de la vie d’un client assuré

Interactions between the events in the life of an insured client

Janosch Ortmann, UQAM

Sorana Froda, UQAM

Juli Atherton, UQAM

Edna Paola Gutiérrez Vega, CIMAT

Carmelle Chango, UQAM

Lajmi Lakhal, Université de Laval

Houyam Dehbi, Beneva

S. Ait Kaci Azzou, Beneva

beneva

Recall goals/objectives

- **Primary goal**

  Use competing risk models to identify subjects with a faster return to work (prioritize the processing of disability cases)

- **Secondary goal**

  Both for pedagogical reasons during the workshop and to compare results from different approaches we considered:

  1) regression
  2) random forests

Overview of presentation

- ✓ Data presentation and descriptive statistics
- ✓ Overview of competing risks theory
- ✓ Regression approach (theory and results)
- ✓ Random forest approach (theory and results)
- ✓ Comparison of methods
- ✓ Future developments
- ✓ Questions?

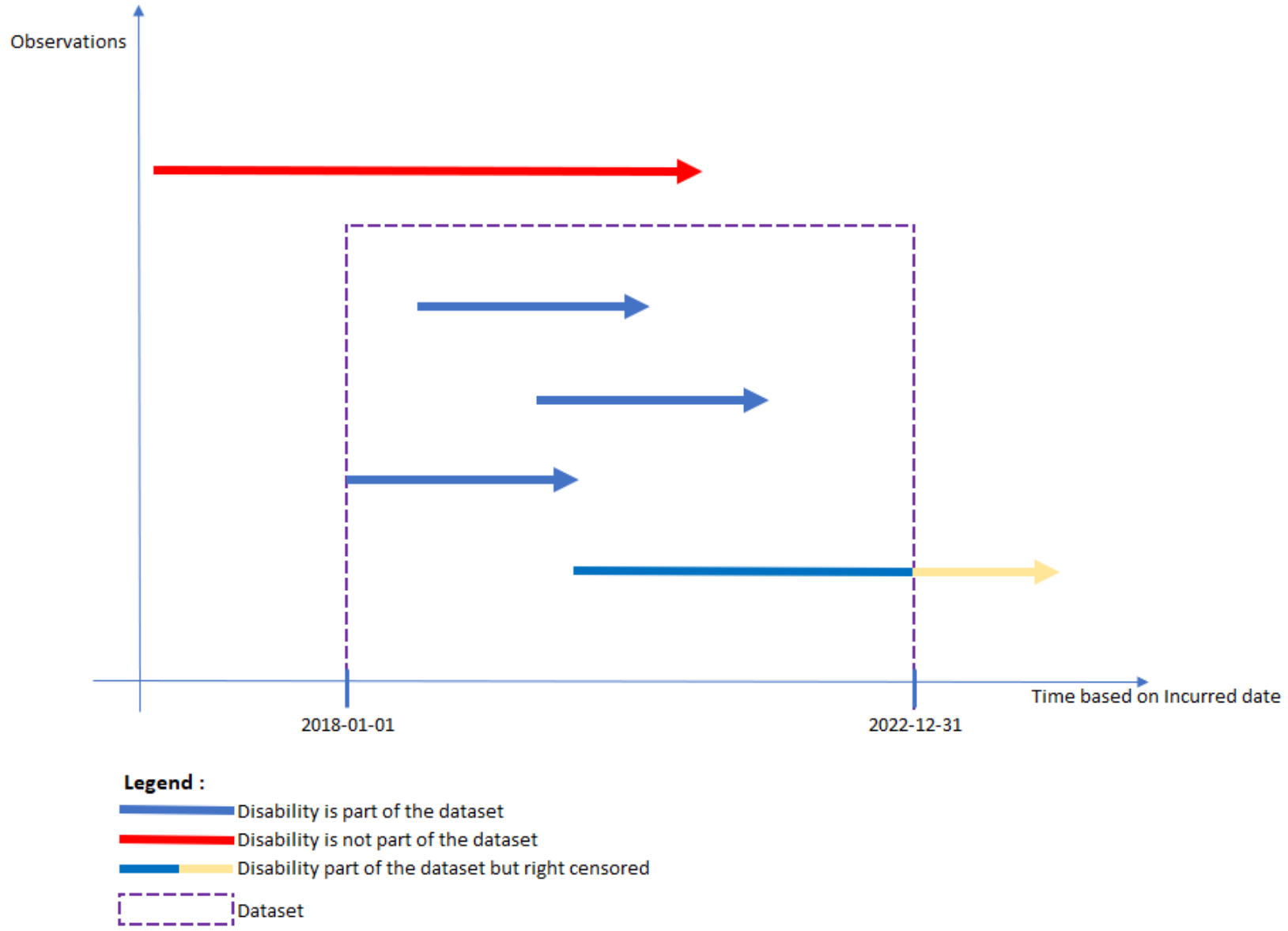
Data description

- **Experience period** : 2018-01-01 to 2022-12-31 (5 years)
- **Date reference** : the experience period is based on the **Incurred Date**
- **Evaluation date** : we assume that the evaluation date is the end of the experience period
- **Population** : all the long-term disabilities with an incurred date between 2018-01-01 and 2022-12-31 inclusively. No conditions on the end of the disability date, the age, the type of insurance contract, the location or other were applied.
- **Data source** : Internal data: Beneva insured only.

Data description

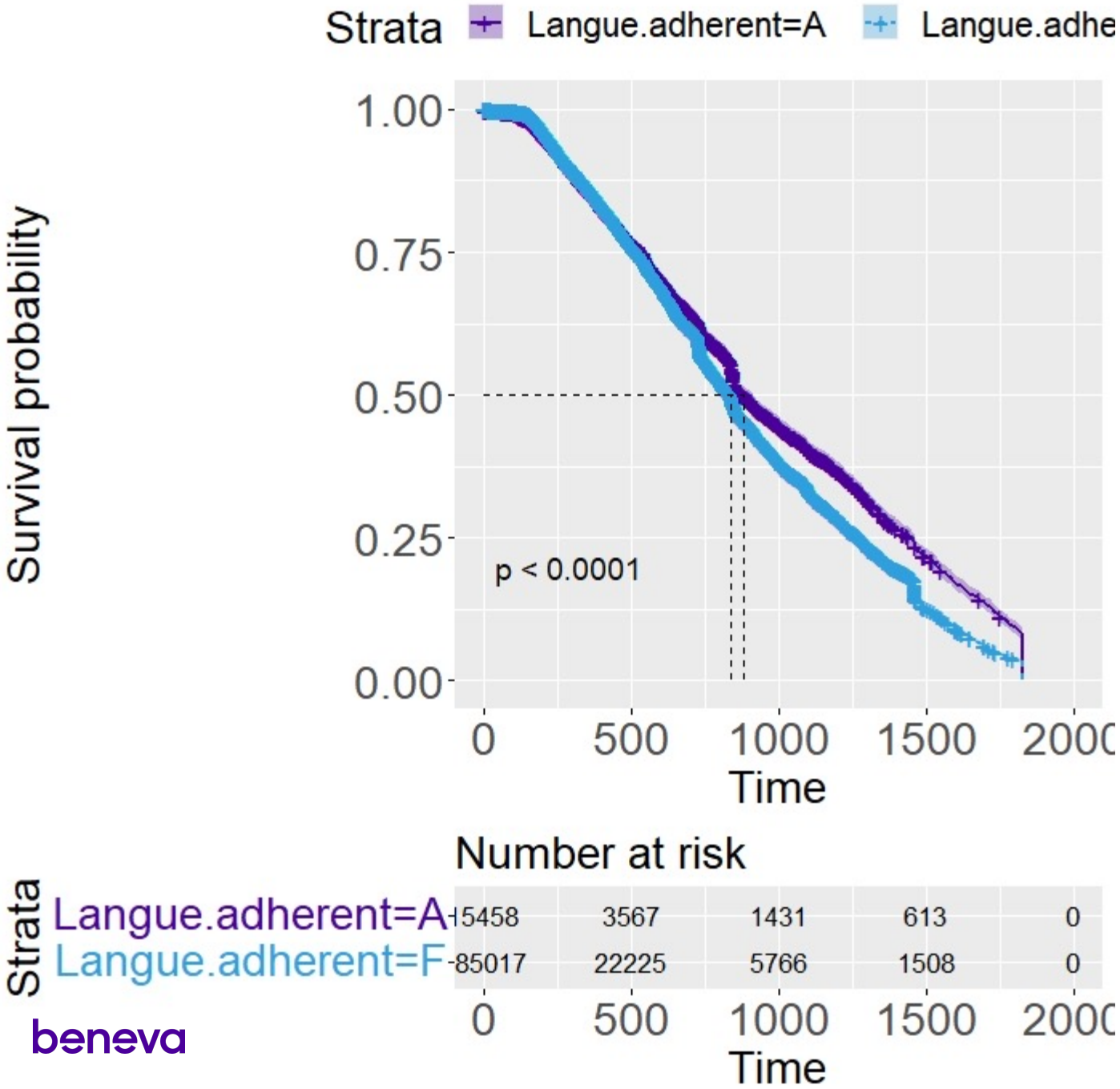
Variable	Description
ID	Unique identifier (for a unique record) based on certificate number and disability file number
Sexe	Gender of the insured
Indicateur.cas.accident	Indicates if the disability is a result of an accident
Indicateur.cas.ITP	Indicates if the disability was stated as total and permanant
Indic.autre.source.revenu	Indicates if the disabled receive disability benefits from other sources than the insurer
Indic.inscription.releve.electronique	Indicates if the participant has an electronic statement subscription
Langue.adherent	Language used by the participant
Salaire.mensuel.brut.au.debut.invalidite	Gross monthly income at the beginning of the disability
Garantie.abregee	Coverage. "Autre" indicates that Beneva provides admisnitrative services only and do no pay any claim.
Code.classe.maladies	Illeness that causes de disability, if there is any.
Etat.civil	Marital status of the participant
Classe.emploi	Job class according to activity level.
Montant.cumulatif.des.prestations	Total claims paid in $ up to evaluation date.
survie	Duration from the incurred date of disability to evaluation date if the insured is still disabled at the evaluation date to the disability end date if the insured is no more disabled at the evaluation date.
Raison.de.fin.invalidite.regroupee	Standardized reasons why a disability would end
Code_RTA	Indicates the geographic location
Age_Debut_Invalidite_groupe	Group of age at the beginning of the disability
recidives_cat	Indicates the number of times the same insured was disabled for what is considered as the same disability (depending on the insurance contract).

Experience description



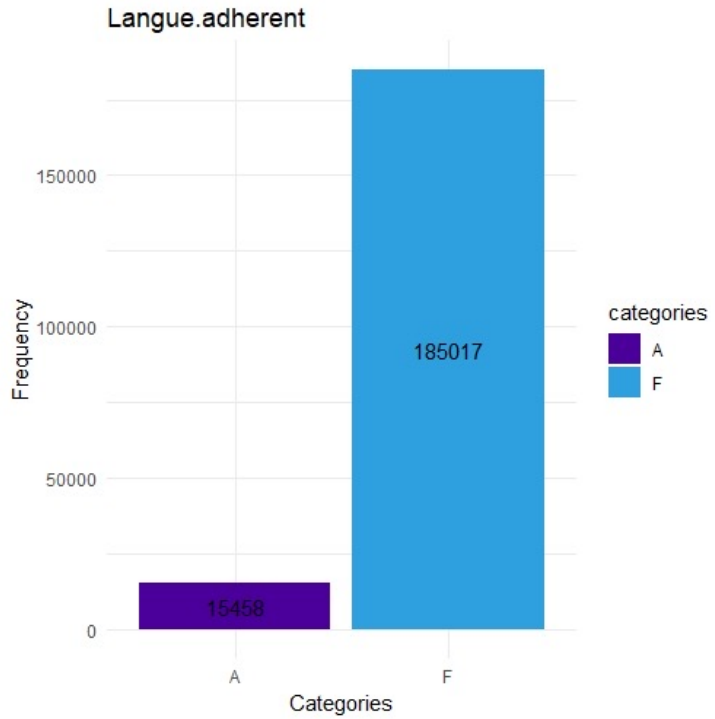
Experience description

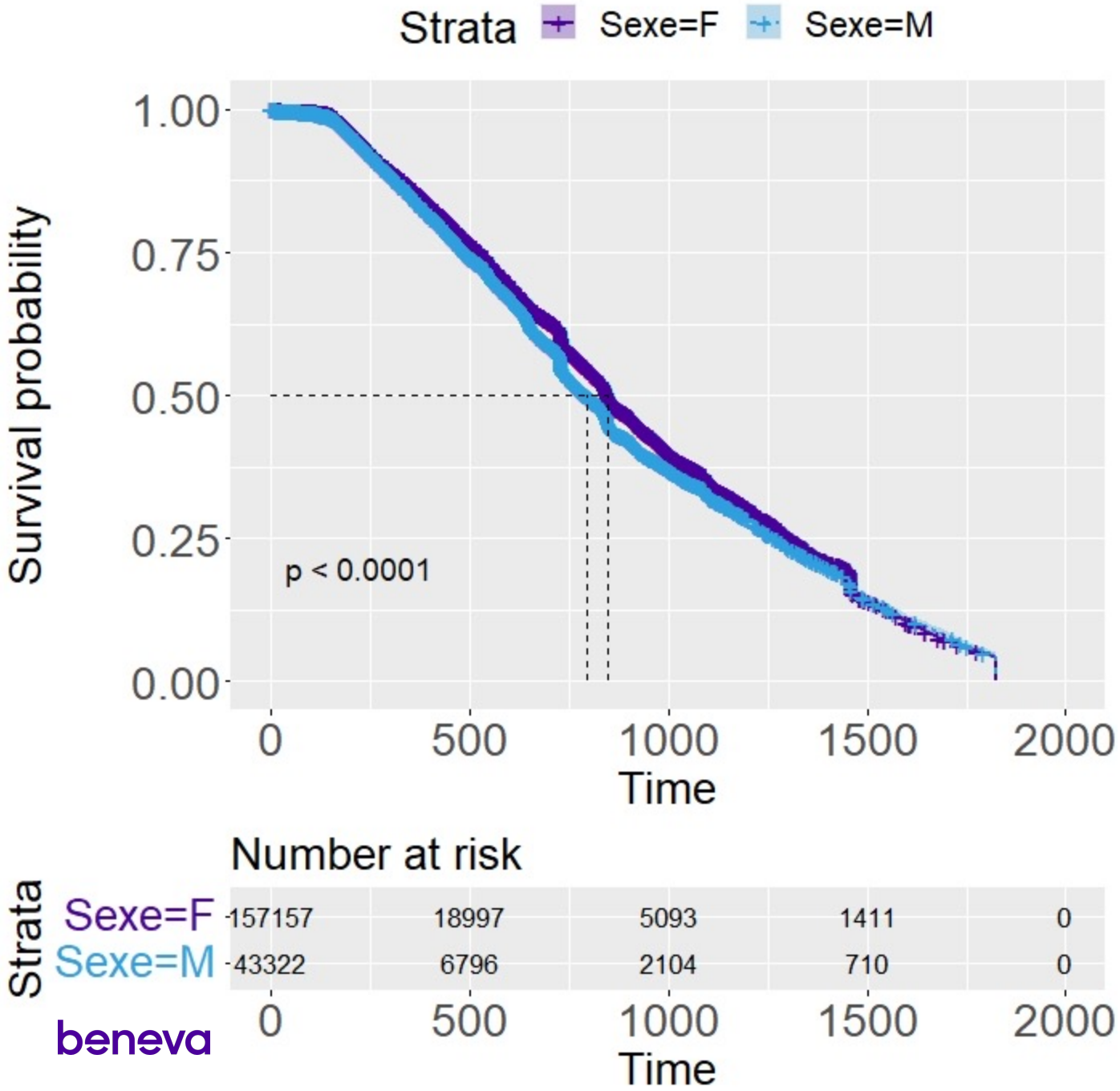
Variable	Description	Statistical Summary
Survie	Duration (days) from the incurred date of disability to evaluation date if the insured is still disabled at the evaluation date to the disability end date if the insured is no more disabled at the evaluation date.	Mean RAT (167.1188) Mean Censurés (664.9083) Mean Décès (516.3672) Mean Règlement Forfaitaire (831.3234) Mean Annulation (780.7828) Mean RRQ/RPC (440.6976) Mean Terminaison (699.7831) Mean Autre (447.5614)
Raison de fin invalidité regroupée	Standardized reasons why a disability would end	RAT (171964, 85.77%) Censuré (20012, 9.9%) Décès (1173, 0.58%) Règlement Forfaitaire (61, ~0%) Annulation (3285, 1.6%) RRQ/RPC (643, 0.32%) Terminaison (1849, 0.92%) Autre (1492, 0.74%) Total (200479)



Covariate:

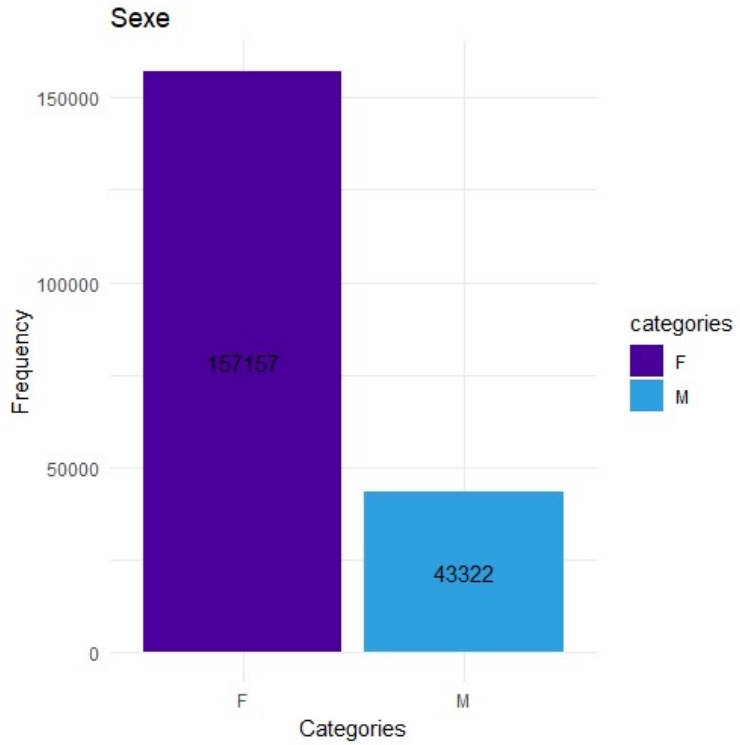
Language used by the participant



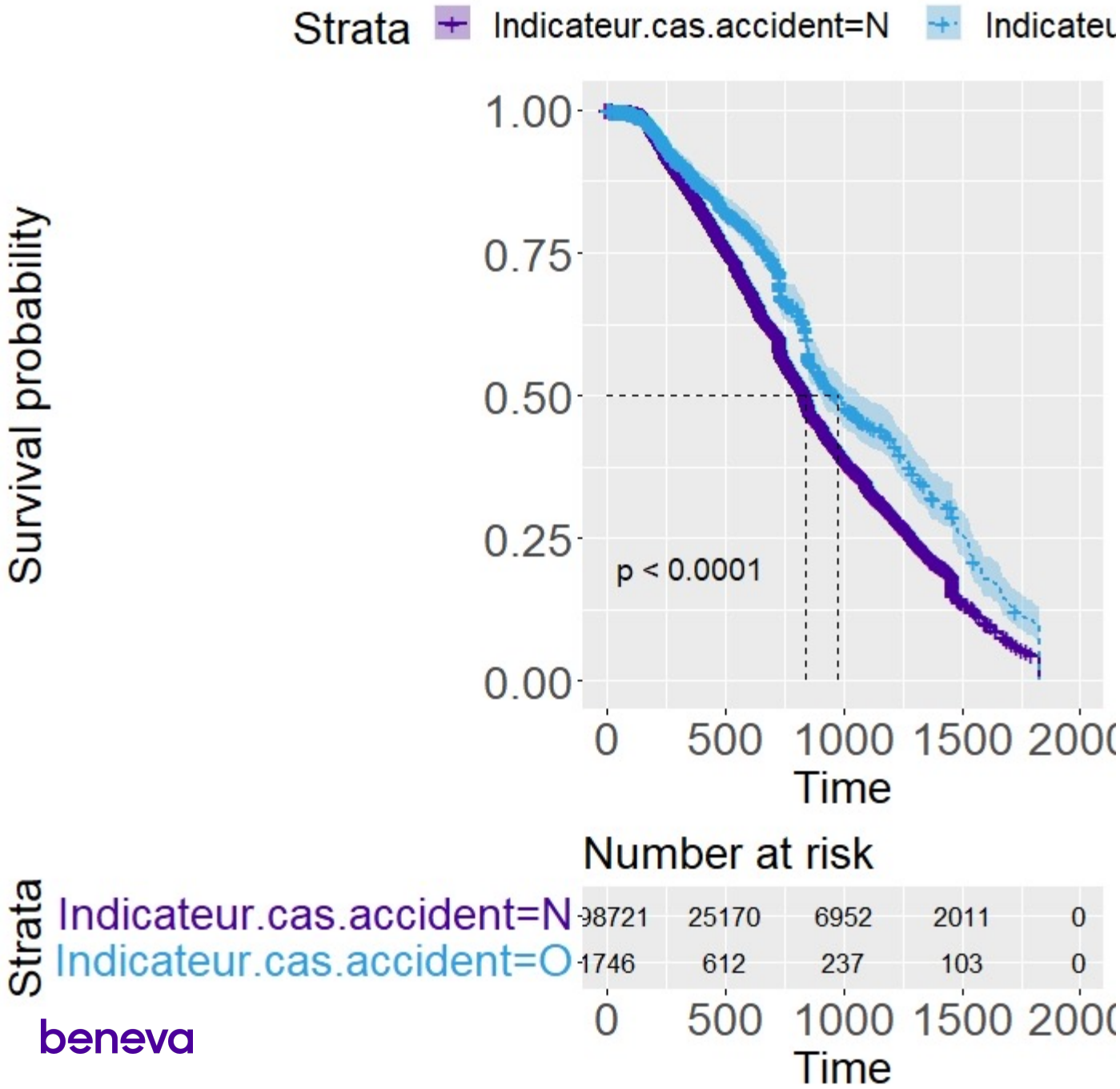


Covariate:

Gender of the insured

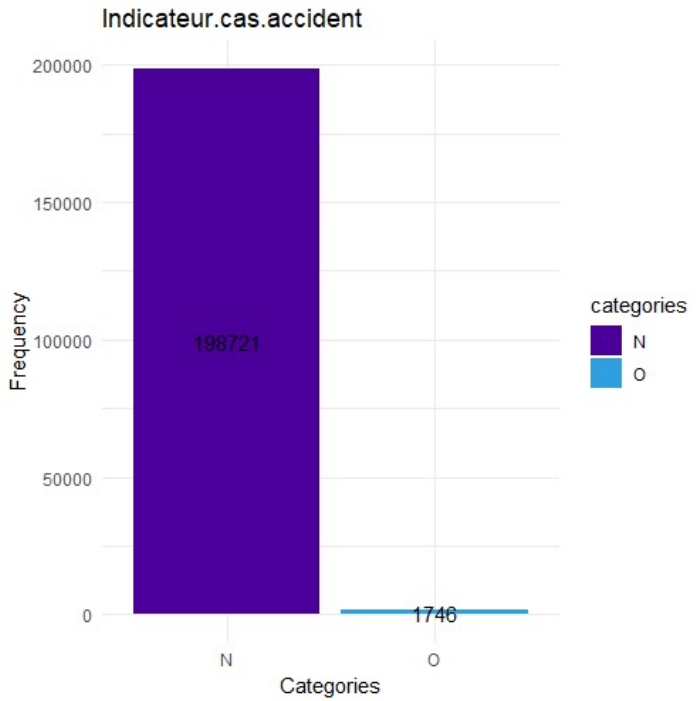


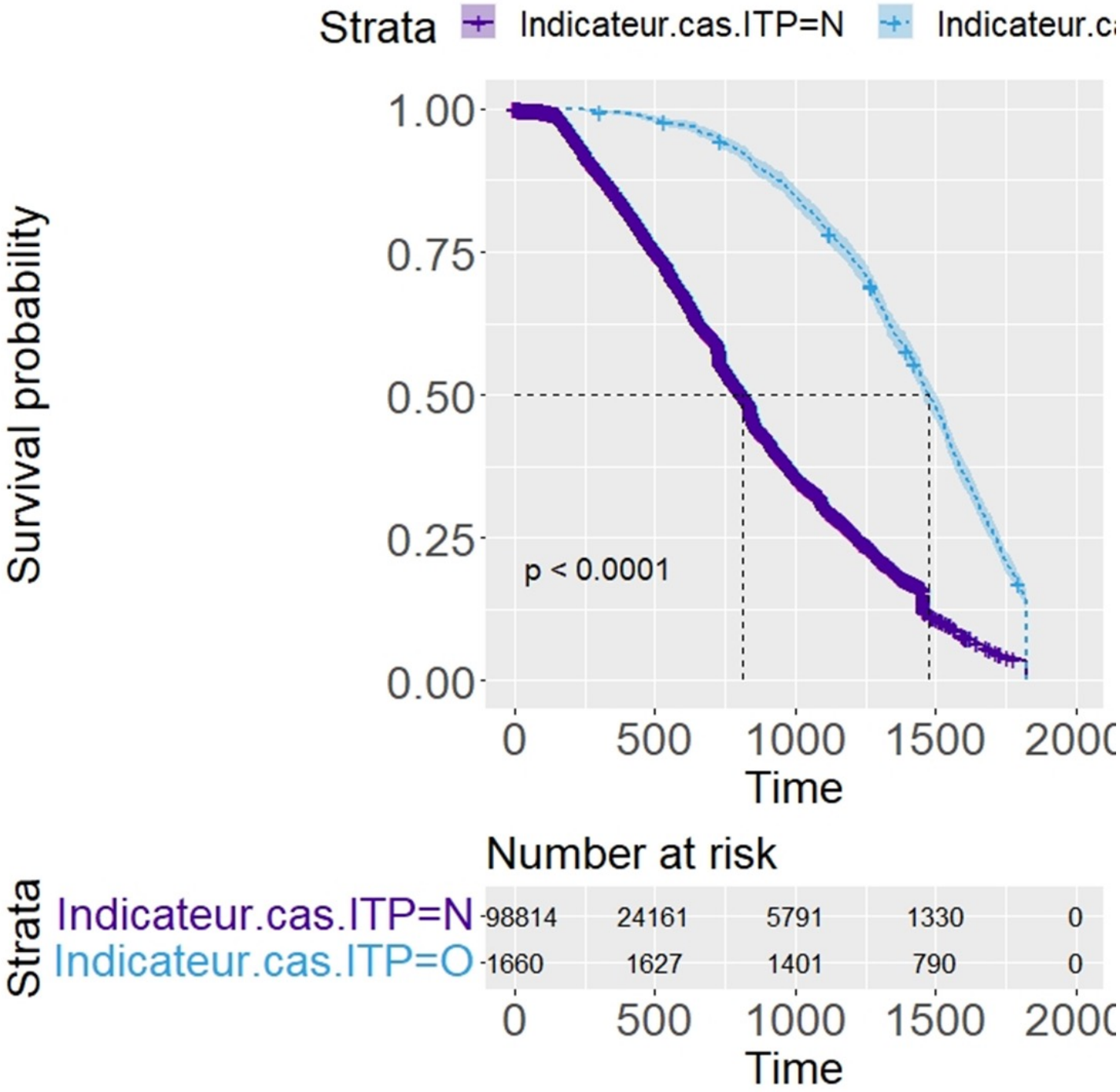
beneva



Covariate:

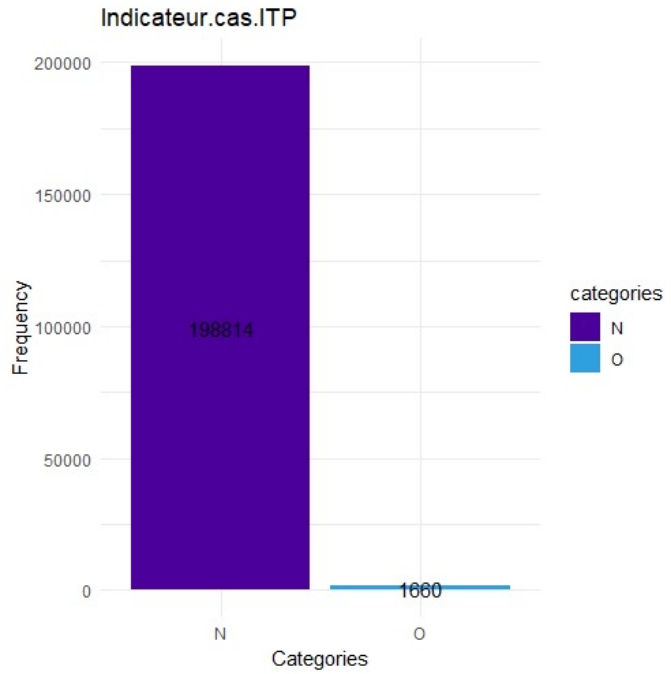
Whether the disability is a result of an accident



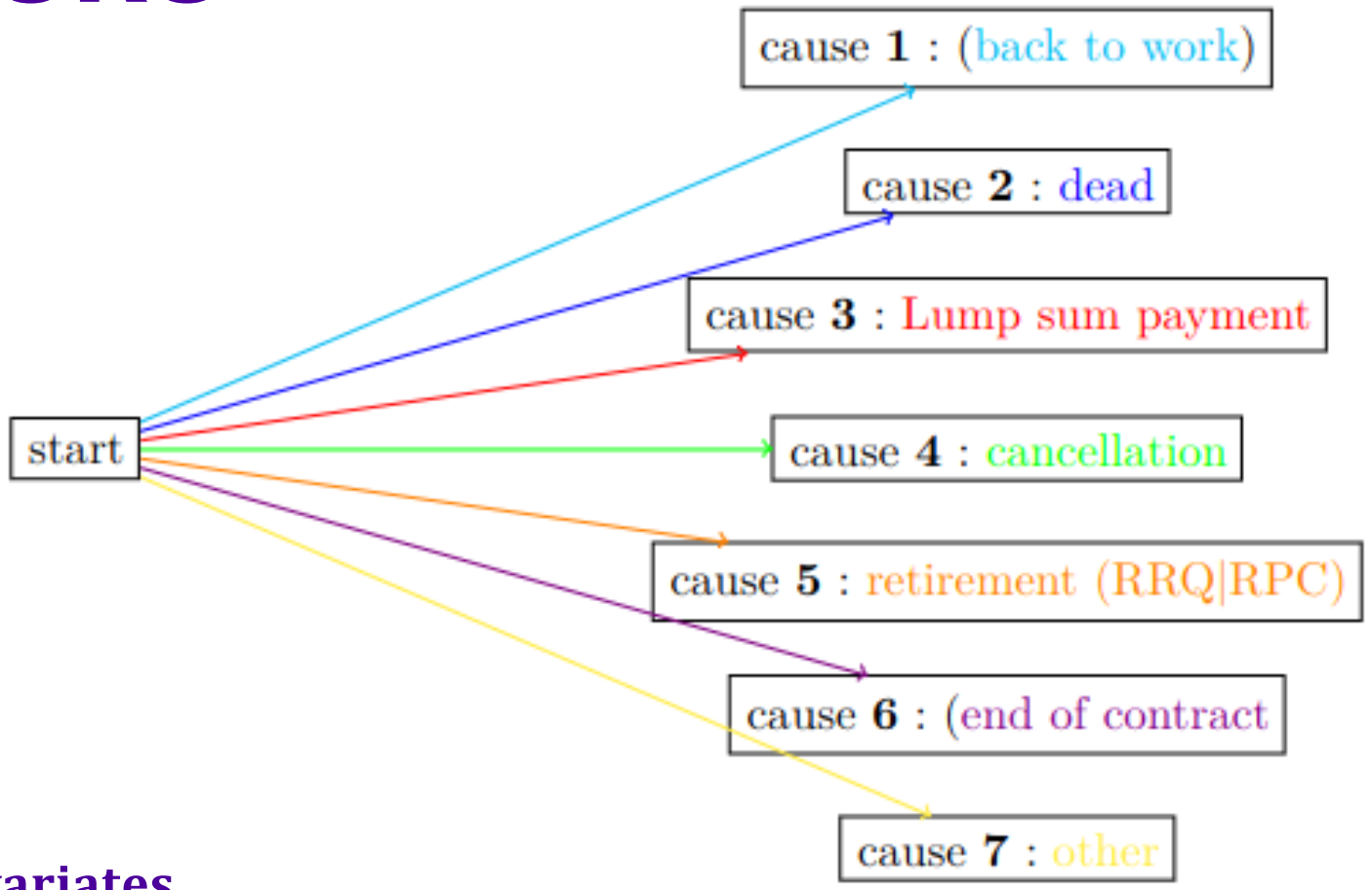


Covariate:

Whether the disability was stated as total and permanent



Competing risks



- $\mathbf{data} = \{(T_i, D_i, Z_i), i = 1, 2, \dots n\}$
- $T_i = \mathbf{observed\ time}$
- $D_i \in \{0, 1, \dots, K\} = \mathbf{set\ of\ cause}$
- $Z_i = (Z_{i1}, \dots, Z_{ip}) = \mathbf{vector\ of\ covariates}$

Multistate model

Hazard function

- $\boldsymbol{\lambda_k(t|z) = \lim_{dt \to 0} \frac{1}{dt} \mathbb{P}(t \leq T < t + dt, D = k | T \geq t, z) \quad k = 1, \dots, K}$

Cumulative hazard function

- $\boldsymbol{\Lambda_k(t|z) = \int_0^t \lambda_k(u|z) du}$

Overall survival function

- $\boldsymbol{S(t|z) = exp\left(-\sum_{k=1}^{K} \Lambda_k(t|z)\right) = \mathbb{P}(T > t|z)}$

Cumulative incidence function

- $\boldsymbol{CIF_k(t|z) = \mathbb{P}(T \leq t, D = k|z) = \int_0^t \lambda_k(u|z) S(u|z) du}$

Hazard function (Cox form – proportional hazard – NOT in random forests)

$\boldsymbol{\lambda_k(t|z) = \lambda_{k0}(t) \cdot exp(\beta_k z)}$

Our regression approach

Proportional hazard (Fine and Gray model)

Hazard function

- $\tilde{\lambda}_k(t|z) = \lim_{dt \to 0} \frac{1}{dt} \mathbb{P}(t \leq T < t + dt, D = k | T \geq t \cup (T \leq t \cap D \neq k), z) \quad k = 1, \ldots, K$

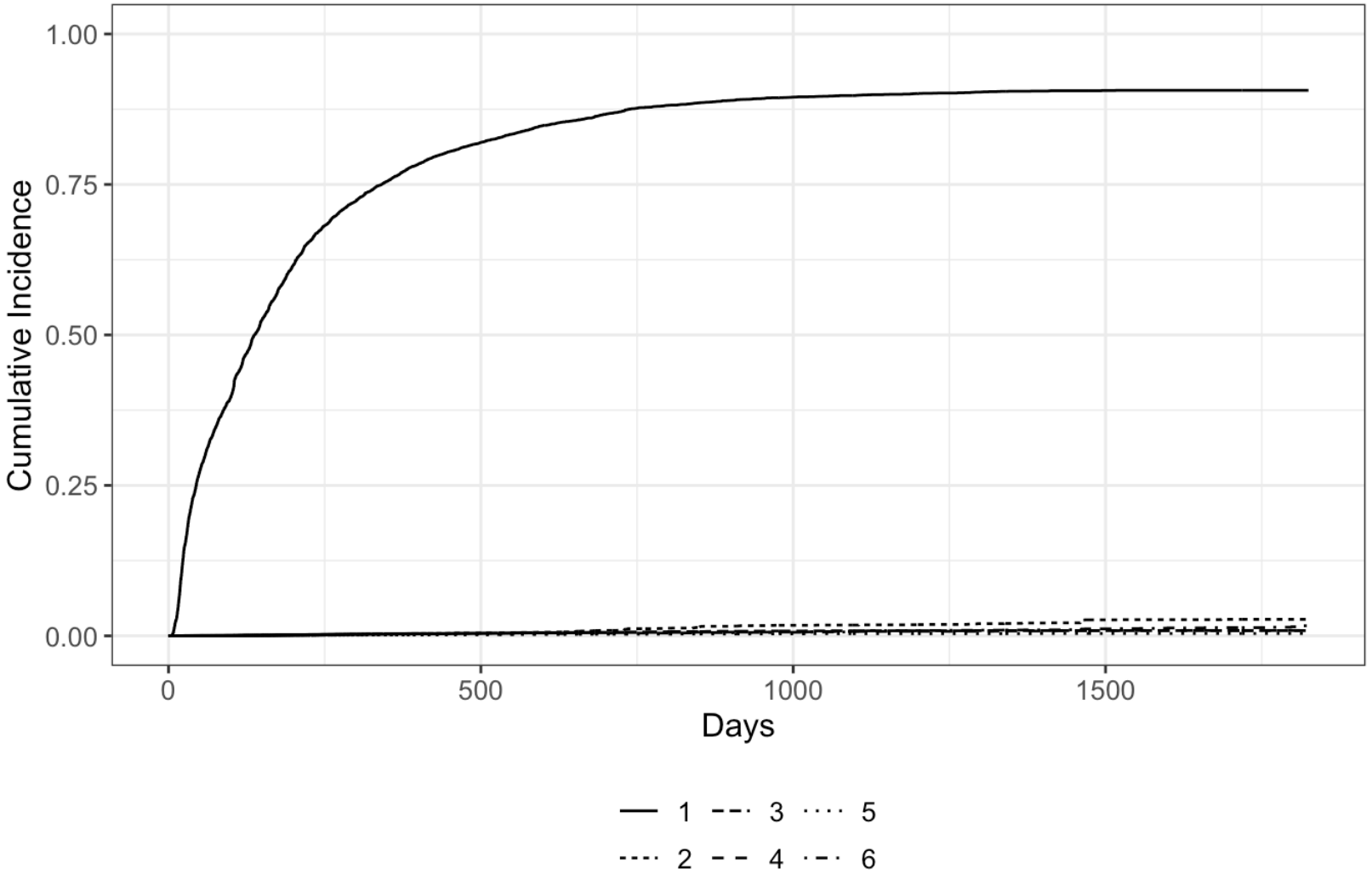
Cumulative incidence function

- $CIF_k(t|z) = 1 - exp\left(-\int_0^t \tilde{\lambda}_k(u|z) du\right)$

Hazard function (Cox form)

$\tilde{\lambda}_k(t|z) = \tilde{\lambda}_{k0}(t) \cdot exp(\beta_k z)$

Most individuals return to work



Results for some covariates

Characteristic	HR[1]	95% CI[1]	p-value
Sexe			
F	—	—	
M	1.11	1.05, 1.17	<0.001
Indicateur.cas.accident			
N	—	—	
O	1.20	0.90, 1.61	0.2
Indicateur.cas.ITP			
N	—	—	
O	0.03	0.01, 0.09	<0.001

Results for some covariates

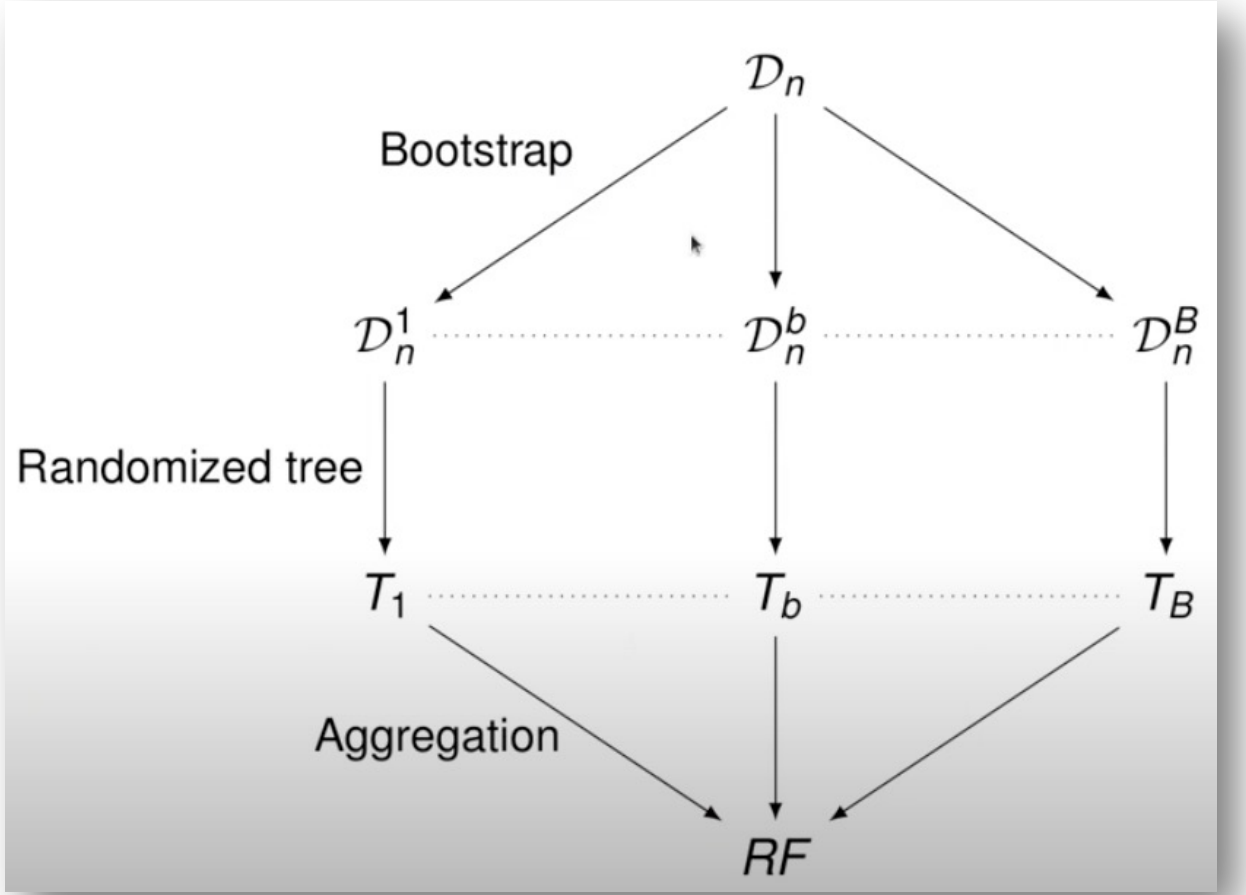
Characteristic	HR_{NC}	HR[1]	95% CI[1]	p-value
Sexe				
F		—	—	
M	1.14	1.11	1.05, 1.17	<0.001
Indicateur.cas.accident				
N		—	—	
O	0.91	1.20	0.90, 1.61	0.2
Indicateur.cas.ITP				
N		—	—	
O	0.43	0.03	0.01, 0.09	<0.001

Significant covariates

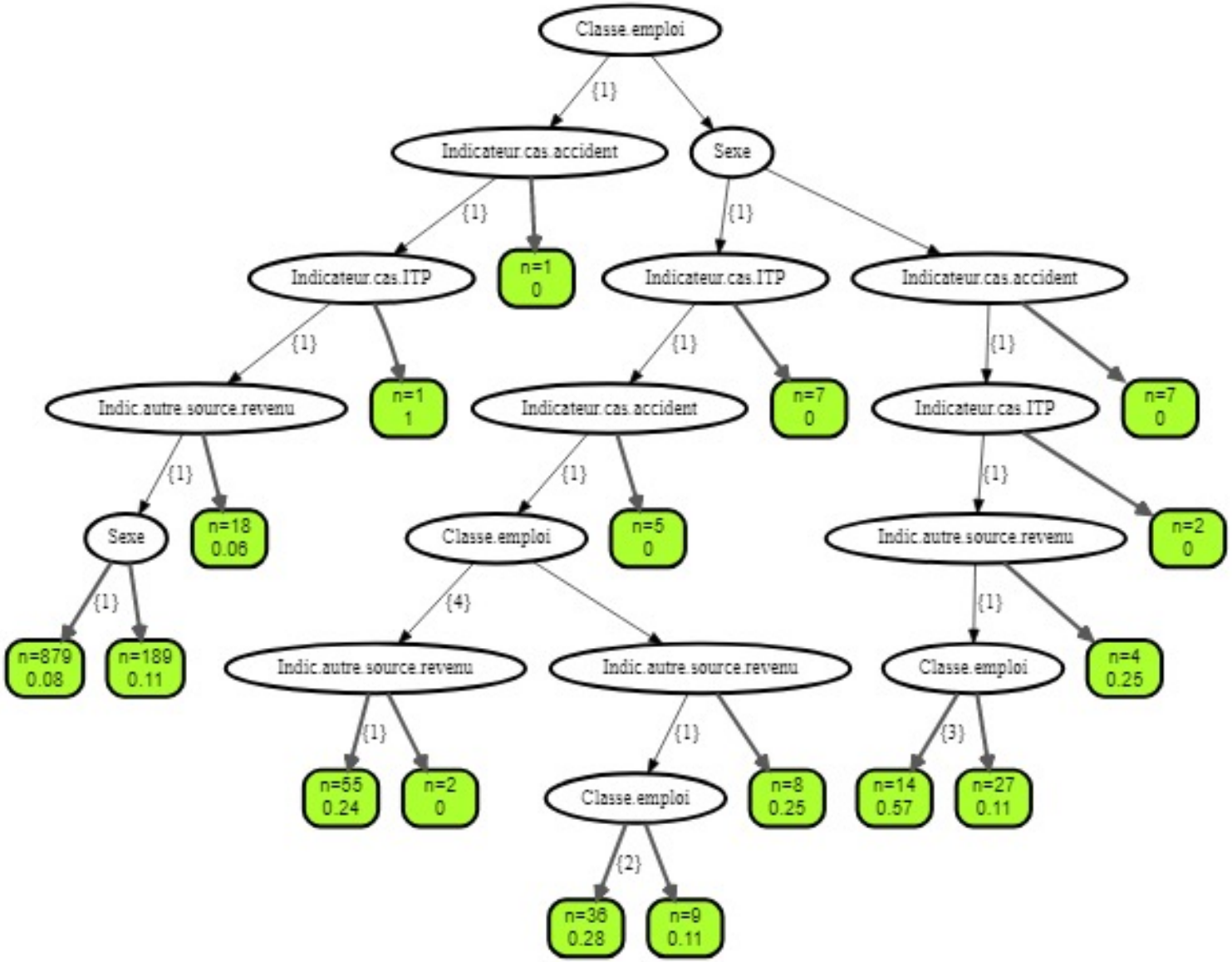
Variable	Description
Sexe	Gender of the insured
Indicateur.cas.ITP	Indicates if the disability was stated as total and permanent
Salaire.mensuel.brut.au.debut.invalidite	Gross monthly income at the beginning of the disability
Code.classe.maladies	Illness that caused the disability, if any.
Etat.civil	Marital status of the participant
Classe.emploi	Job class according to activity level.
Montant.cumulatif.des.prestations	Total claims paid in $ up to evaluation date.
Age_Debut_Invalidite_groupe	Group of age at the beginning of the disability

Random forests for competing risks

Introduction

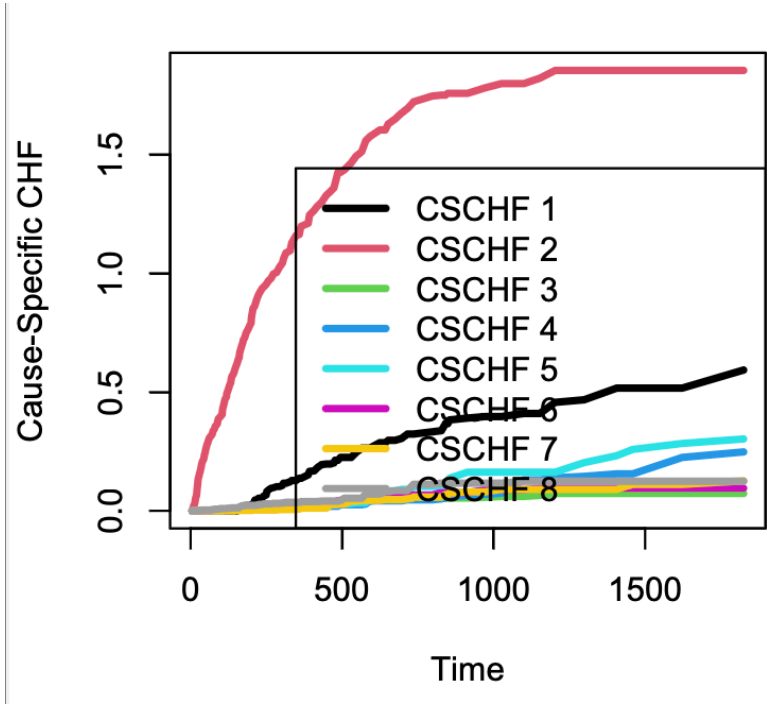


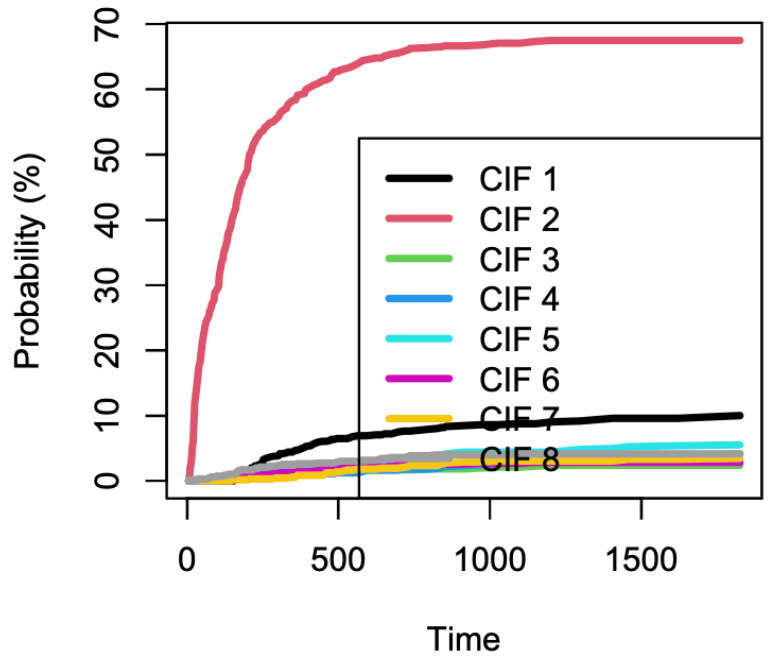
Introduction



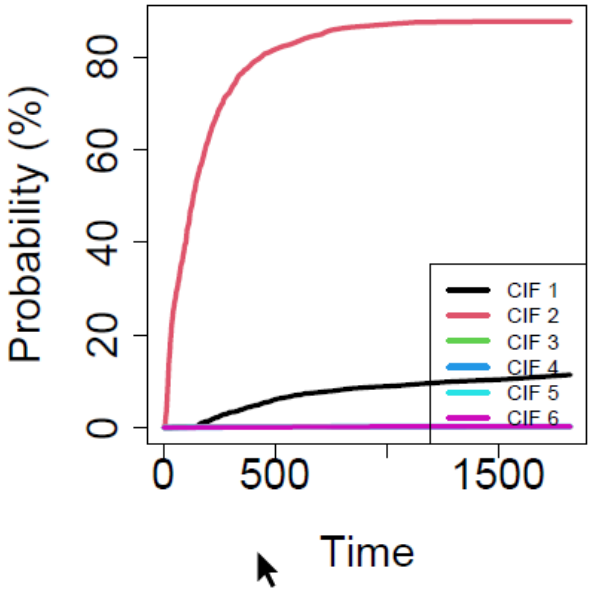
Random forest results

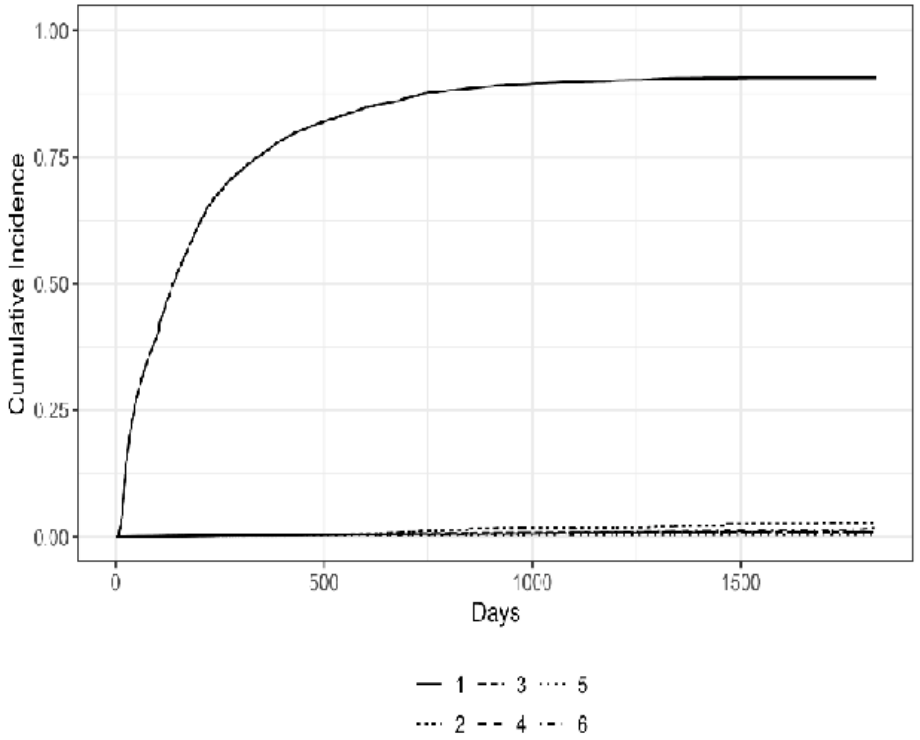
- **Preliminary results using VIMP (variable importance) and a subsample of the data suggests the variables: indicateur autre source revenu, langue adherent, salaire mensuel brut au début invalidité and age début invalidité groupe**





Compare and summarize results





Future work (report)

- Investigate the predictability and explainability for the two classes of methods
- Redo the survival random forest with competing risks analyses using the whole dataset.
- Explore other packages using big data (largeRCRF for example)
- Go deeply to the theoretical aspects
- Evaluate and compare model performance for both methods using metrics such as (C-index, Brier, …)

Merci / Thanks

Questions ?

beneva

References

beneva

Beneva

75 ans d’histoire

- **Compagnie d’assurance**
- **Né du regroupement La Capitale et de SSQ Assurance**
- **Assurances auto, habitation, vie, entreprise et produits financiers**

75 years in the making

- **Insurance company**
- **We are La Capitale and SSQ Insurance coming together as one**
- **Auto, home, life, commercial insurances and financial products**

$26,8 billion
in assets
en actif

3,5 million
clients and members
de membres et clients

5 000
employees
employés

Detailed data description

Variable	Description	Statistical Summary
ID	Unique identifier (for a unique record) based on certificate number and disability file number	
Sexe	Gender of the insured	M (43322, 21.6%) F (157157, 78.3%) Total (200479)
Indicateur cas accident	Indicates if the disability is a result of an accident	Y (1746, 0.0087%) N (198721, 99.1%) Total (200467)
Indicateur cas ITP	Indicates if the disability was stated as total and permanant	Y (1660, 0.0082%) N (198814, 99.17%) Total (200474)
Indicateur autre source revenu	Indicates if the disabled receive disability benefits from other sources than the insurer	Y (6031, 0.03%) N (194448, 96.9%) Total (200479)

Detailed data description

Variable	Description	Statistical Summary
Indicateur inscription relevé électronique	Indicates if the participant has an electronic statement subscription	Y (83314, 41.4%) N (117165, 58.4%) Total (200479)
Langue adherent	Language used by the participant	E (15458, 7.7%) F (185017, 92.28%) 9 (4, ~0%) Total (200479)
Salaire mensuel brut au début invalidité	Gross monthly income at the beginning of the disability	Mean (4530.603), Median (3909), Min (0), Max (833338) sd (3301.257)
Garantie abrégée	Coverage. "Autre" indicates that Beneva provides administrative services only and does not pay any claim.	Autre (172018, 85.8%), FG (115, ~0%), LD (27548, 13.74%), LD add (798, 0.003%) Total (200479)
Code Classe maladie	Illness that causes the disability, if there is any.	1 (1052), 2(111), 3 (431), 4 (643), 5 (218), 6 (1161), 7 (158), 8 (768), 9 (1954), 10 (205), 11 (8041), 12 (504), 13 (10526), 14 (84), 15 (256), 99 (174352) Total (200465)

Detailed data description

Variable	Description	Statistical Summary
Etat civil	Marital status of the participant	Autre (152753, 76.2%) Célibataire (31993, 15.96%) Conjoint de fait (3867, 1.92%) Divorcé (1312, 0.65%) Inconnu (32, 0.016%) Marié (9856, 4.9%) Séparé (449, 0.22%) Veuf (144, 0.07%) Total (200406)
Classe emploi	Job class according to activity level.	Inconnu (169553, 84.6%) Léger (9943, 4.9%) Lourd (7134, 3.5%) Sédentaire (13849, 6.9%) Total (200479)
Montant cumulatif des présentations	Total claims paid in $ up to evaluation date.	Mean (154.46), Median (0), Min (0), Max (523800), sd(3614.831)

Detailed data description

Variable	Description	Statistical Summary
Survie	Duration from the incurred date of disability to evaluation date if the insured is still disabled at the evaluation date to the disability end date if the insured is no more disabled at the evaluation date.	Mean RAT (167.1188) Mean Censurés (664.9083) Mean Décès (516.3672) Mean Règlement Forfaitaire (831.3234) Mean Annulation (780.7828) Mean RRQ/RPC (440.6976) Mean Terminaison (699.7831) Mean Autre (447.5614)
Raison de fin invalidité regroupée	Standardized reasons why a disability would end	RAT (171964, 85.77%) Censuré (20012, 9.9%) Décès (1173, 0.58%) Règlement Forfaitaire (61, ~0%) Annulation (3285, 1.6%) RRQ/RPC (643, 0.32%) Terminaison (1849, 0.92%) Autre (1492, 0.74%) Total (200479)

Detailed data description

Variable	Description	Statistical Summary
Code RTA	Indicates the geographic location	1221 categories
Age début invalidité groupe	Group of age at the beginning of the disability	55+ (36971, 18.44%) [47,55[(42869, 21.3%) [40,47[(38260, 19%) [32,40[(42060, 20.9%) [15,32[(40319, 20.1%) Total (200479)
Récidives cat	Indicates the number of times the same insured was disabled for what is considered as the same disability (depending on the insurance contract).	0 (199418, 99.47%) 1 (957, 0.47%) 2+ (104, 0.05%) Total (200479)