

# SNSF Report

*Chiara Gilardi, Leslie O’Bray, Carla Schärer and Tommaso Portaluri*

*5 April 2018*

## Introduction

## Data Description

We have three data sets: Applications, External reviewers and Internal referees.

### Applications

- AmountRequested: Rounded to the next 10k CHF
- AmountGranted: Rounded to the next 10k CHF
- IsApproved: 1 if the application was approved, 0 if it was rejected
- GradeFinal: Comparative ranking of the application as determined by the evaluation body (the division of the National Research Council). A: “belongs to the 10% best percent”; AB: “10% are worse, 75% are better”; B: “50% are worse, 25% are better”; BC: “25% are worse, 50% are better”; C: “10% are worse, 75% are better”; D: “90% of the applications are better”
- Division: Evaluation Body in which the application was evaluated. Division 1 evaluates Social Sciences and Humanities; Division 2 Mathematics, Natural Sciences and Engineering; Division 3 Biology and Medicine
- MainDiscipline: as chosen by the applicant from the SNSF discipline list
- MainDisciplineLevel2: category in the SNF discipline list grouping disciplines into fields of research
- CallTitle: Call for proposals under which the application was submitted. Applications from the same Call are evaluated together, i.e. in competition to each other
- CallEndDate: Submission deadline of the Call
- ResponsibleApplicantAcademicAgeAtSubmission: Years since the applicant’s PhD at time of submission; data only available since mid 2016
- ResponsibleApplicantAgeAtSubmission: Biological age of the applicant at time of submission; data only available since mid 2016
- ResponsibleApplicantProfessorshipType: employment situation of the applicant at time of submission; data only available since mid 2016
- Gender: of the main applicant
- NationalityIsoCode: Nationality of the main applicant
- IsHasPreviousProjectRequested: 0 if it is the applicant’s first application at the SNSF, 1 if not
- InstType: Type of institution where the applicant is employed
- IsContinuation: 1 if the project is a thematic continuation of a previously approved project, 0 if not
- ProjectID: Anonymized identifier of the application

## Referee Grades

- Question: Evaluation criterion
- QuestionRating: The (co-)referee's assessment of the evaluation criterion
- OverallRanking: The (co-)referee's overall comparative ranking of the application. A: "belongs to the 10% best percent"; same scale as the GradeFinal
- RefereeRole: Some applications have one referee evaluation, some have two. The role indicates who was the primary and who was the secondary referee (also called co-referee)
- RefereeGender
- IDs: Anonymized identifiers of the application, the referee and the evaluation by the referee

## Reviews

- Question: Evaluation criterion
- QuestionRating: The external reviewer's assessment of the evaluation criterion
- OverallGrade: The external reviewer's overall assessment of the application
- SourcePerson: Who suggested the reviewer?
- Gender
- Country: where the reviewer is located. Not always known
- EmailEnding: ending of the reviewer's email address. Might be used as an approximation of the country where the reviewer is located in cases where this data is missing
- IDs: Anonymized identifiers of the application, the reviewer and the review

## Cleaning the Data

We decide to work with only complete applications, i.e. project for which we have information from the three data sets.

As we have only information from reviewers and referees since 2016, we are considering applications only from that year.

Specific to each data set this are the detailed considerations:

## Applications

We decide to consider only the MainDiscipline2 because for MainDiscipline we have 118 levels, while for the other only 21.

There is one application for which we do not know the gender of the applicant, and therefore we decided to omit that observation from the analysis.

We will also not consider the variables "CallTitle", "Professorship", "AcademicAge". The first one, because we consider it has nothing to add to the model. The two last, due to the fact that there are a considerable number of NA's on those variables.

```
##
##   Assistant professor with tenure track
##                                     102
## Assistant professor without tenure track
##                                     54
##               Associate professor
##                                     237
##               Full professor
##                                     512
```

```
## Honorary professor or Titular professor
## 77
## None
## 430
## Professor at UAS / UTE
## 74
## Visiting professor
## 4
## <NA>
## 20037
## [1] 0.9361267
```

## Internal Referees

There were 22 observations (1 for the time frame we are dealing with) for which only demographic information was available, no grades were given. We decide to omit those observations.

Also we decide to not consider the Referee role as a variable in our model, as the majority of the evaluations has only one referee.

```
##
## Applicant Explicit inclusion Recusal
## 2 20 8
## Referee Second referee <NA>
## 15766 870 1276
```

## External Reviewers

Reviewers always have the option to choose “not considered” or “0” when reviewing an application. Some might be by mistake, others might have a conflict of interest, or they might be very ambivalent about the project. Therefore, we did not considered observations with this grades. One of the questions evaluated in the applications is “Broader impact (forms part of the assessment of scientific relevance, originality and topicality)”.

For the time frame we are considering, in all the applications this grade was NA. Hence, we omit this variable from our model.

## Exploratory Analysis

### Analysis

#### Internal Review regression

```
##
## Call:
## glm(formula = external_regression_data$IsApproved ~ . - (ProjectID),
## family = "binomial", data = external_regression_data)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -1.5800 -1.1912 0.8749 1.0813 1.6736
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.232837   0.439763   0.529 0.596485
## ApplicantTrack -0.222089   0.055872  -3.975 7.04e-05 ***
## ScientificRelevance -0.099758  0.060181  -1.658 0.097394 .
## Suitability     0.036162   0.054736   0.661 0.508831
## OverallGrade     0.028024   0.062485   0.448 0.653801
## PercentFemale   -0.613661   0.216469  -2.835 0.004585 **
## Age             0.010733   0.006313   1.700 0.089121 .
## Genderm         0.135412   0.125794   1.076 0.281719
## DivisionDiv 2    0.502163   0.132672   3.785 0.000154 ***
## DivisionDiv 3    0.359610   0.127477   2.821 0.004788 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 2243.9  on 1622  degrees of freedom
## Residual deviance: 2176.9  on 1613  degrees of freedom
## AIC: 2196.9
##
## Number of Fisher Scoring iterations: 4
```

## External review analysis

```
##
## Call:
## glm(formula = external_regression_data$IsApproved ~ . - (ProjectID),
##      family = "binomial", data = external_regression_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5800  -1.1912   0.8749   1.0813   1.6736
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.232837   0.439763   0.529 0.596485
## ApplicantTrack -0.222089   0.055872  -3.975 7.04e-05 ***
## ScientificRelevance -0.099758  0.060181  -1.658 0.097394 .
## Suitability     0.036162   0.054736   0.661 0.508831
## OverallGrade     0.028024   0.062485   0.448 0.653801
## PercentFemale   -0.613661   0.216469  -2.835 0.004585 **
## Age             0.010733   0.006313   1.700 0.089121 .
## Genderm         0.135412   0.125794   1.076 0.281719
## DivisionDiv 2    0.502163   0.132672   3.785 0.000154 ***
## DivisionDiv 3    0.359610   0.127477   2.821 0.004788 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 2243.9  on 1622  degrees of freedom
## Residual deviance: 2176.9  on 1613  degrees of freedom
```

```
## AIC: 2196.9
##
## Number of Fisher Scoring iterations: 4
```

## Results

## Conclusion