

Milestone Data 101

Xilin Chen

2020-01-29

About this Document

This document explores some basic quality stats for the Milestone data received on Jan 2020.

Observations: 370,347

Variables: 25

\$ ScheduleWindowDescription	<chr>	"2014-2015 ACGME Mid-Year Milestone Evaluations", "2014-2015 ACGME Mi
\$ SurveyDefinitionID	<chr>	"85", "85", "85", "85", "85", "85", "85", "85", "85", "85", "85", "85", "85"
\$ SurveyName	<chr>	"2014-2015 ACGME Milestone Evaluations - Surgery", "2014-2015 ACGME M
\$ AcademicYear	<chr>	"2014", "2014", "2014", "2014", "2014", "2014", "2014", "2014", "2014"
\$ ProgramId	<chr>	"6665", "6665", "6665", "6665", "6665", "6665", "6665", "6665", "6665"
\$ SpecialtyCode	<chr>	"440", "440", "440", "440", "440", "440", "440", "440", "440", "440", "
\$ SpecialtyName	<chr>	"Surgery", "Surgery", "Surgery", "Surgery", "Surgery", "Surgery", "Surgery", "Sur
\$ SponsorID	<chr>	"12181", "12181", "12181", "12181", "12181", "12181", "12181", "12181", "12181"
\$ PersonID	<chr>	"745633", "745633", "745633", "745633", "745633", "745633", "745633", "745633", "
\$ NPI	<chr>	"1881039279", "1881039279", "1881039279", "1881039279", "1881039279", "
\$ Last4SSN	<chr>	"7686", "7686", "7686", "7686", "7686", "7686", "7686", "7686", "7686", "
\$ BirthDate	<chr>	"01JAN1984", "01JAN1984", "01JAN1984", "01JAN1984", "01JAN1984", "01J.
\$ MedicalSchoolID	<chr>	"18838", "18838", "18838", "18838", "18838", "18838", "18838", "18838", "18838"
\$ DegreeDate	<chr>	"01JUL2012", "01JUL2012", "01JUL2012", "01JUL2012", "01JUL2012", "01J.
\$ residentyear	<chr>	"2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "2", "
\$ PositionType	<chr>	"Categorical", "Categorical", "Categorical", "Categorical", "Categori
\$ ResidentStatus	<chr>	"Active Full time", "Active Full time", "Active Full time", "Active F
\$ StartDate	<chr>	"07/01/2013", "07/01/2013", "07/01/2013", "07/01/2013", "07/01/2013", "
\$ ExpectedCompletionDate	<chr>	"06/30/2018", "06/30/2018", "06/30/2018", "06/30/2018", "06/30/2018", "
\$ QuestionID	<chr>	"9838", "9840", "9842", "9845", "9847", "9850", "9861", "9853", "9863"
\$ QuestionKey	<chr>	"comp1_PC_Q1", "comp1_PC_Q2", "comp1_PC_Q3", "comp2_MK_Q1", "comp2_MK
\$ ReportCategory	<chr>	"Patient Care", "Patient Care", "Patient Care", "Medical Knowledge", "
\$ QuestionText	<chr>	"Care For Diseases and Conditions (CDC) - PATIENT CARE (PC1)", "Care I
\$ IntResponseValue	<chr>	"5", "5", "4", "5", "5", "5", "5", "5", "5", "5", "5", "5", "5", "5", "5", "
\$ ResponseText	<chr>	"<p><u>Level 2</u></p> <p>This resident accurately diagnoses

Data

Person ID and NPI values missingness

Table 1: NPI and PersonID

n_tot_person	missing_npi	perc_missing_npi	n_missing_personID
4744	1648	0.35	1

Last4SSN

\begin{table}[!h]

\caption{Example of non_numeric SSN}

PersonID	NPI	Last4SSN
745648	1700129772	*****7544
745655	1407290844	*****7083
745674	1295178697	*****5251
745677	NA	*****7922
749756	1265877781	*****1161
749763	1215372768	*****9640

\end{table}

Table 2: one person has less than 4 digits of SSN

PersonID	NPI	Last4SSN
982347	NA	967

Date

some Dates are not dates

```
# A tibble: 6 x 3
  PersonID NPI      BirthDate
  <chr>    <chr>    <chr>
1 745648  1700129772 *****
2 745655  1407290844 *****
3 745674  1295178697 *****
4 745677  <NA>      *****
5 749756  1265877781 *****
6 749763  1215372768 *****

# A tibble: 0 x 3
# ... with 3 variables: PersonID <chr>, NPI <chr>, StartDate <chr>

# A tibble: 6 x 3
  PersonID NPI      DegreeDate
  <chr>    <chr>    <chr>
1 745648  1700129772 *****
2 745655  1407290844 *****
3 745674  1295178697 *****
4 745677  <NA>      *****
5 749756  1265877781 *****
6 749763  1215372768 *****

# A tibble: 0 x 3
# ... with 3 variables: PersonID <chr>, NPI <chr>, ExpectedCompletionDate <chr>
```