# Introduction

**Motivation:**
- Electronic Health Record
  - automatic extract
  - important information
  - Data-driven quality measures
- Free-text form

**Objective:**
- We laid out an annotation framework and evaluated it in a case study where quality measures were extracted from colonoscopy reports.
- In this case study, we chose a configuration which can be applied to multiple domains without medical and natural language processing (NLP) expertise but still has reasonable results.

---

### #1: Annotation task

Identifying the problem we are working on (usually done by domain experts/researchers)

**A formal representation is needed in this step for both applying automatic methods and using human annotators.**

For example: 
- Record yes if physician stopped the procedure early due to too much stool, patient discomfort, etc.
- **Automatic methods** is difficult to implement the case etc in a system
- **Human annotators** annotators will be confused about the case etc.

*This step is domain-independent.*

---

### #2: Objective function

Identifying the information we want to annotate. We have 2 possible objective functions in this project

Choosing:

1. **Indicators**
   - did not give good results. Indicators are more abstract (than Variables) and thus difficult to automatically predict given the simple annotating features used in this project.

2. **Variables**
   - gave interesting results, which are reported in Table 2. After extracted values of Variables, we can easily convert them into Indicators by deterministic rules.

Single concept variables are better choice for our framework configuration.

---

### #3: Annotating features

Annotating features are representations of input data. We tried bag of words features

1. They are easy to apply to new domains (which is in our framework configuration)
2. They can be applied quickly to evaluate different approaches
3. They empirically give reasonable results.

---

### #4: Annotating methods

We tried different kinds of method:

1. Simple method: SimpleSQ which is a text searching system
2. Rule-based method: C-QUAL which is proposed by Harkema et al. (2013). This system is considered as the baseline in this project
3. Generic machine learning (ML) methods: Support vector machine (SVM) and Decision tree J48. These methods are general and domain-independent.

As shown in Table 2, the simple method worked well in many cases. Meanwhile, ML methods improved some cases where the simple method did not work well.

---

### #5: Skewed data

Data in clinical domains is often skewed. We tried 2 popular re-sampling methods, which are generic and domain-independent, to handle skewed data (shown in Table 1), i.e. over-sampling and down-sampling.

As shown in Figure 1 and Figure 2, re-sampling methods only helped minority classification and degraded majority class classification. Moreover, re-sampling methods only help when the data distribution is not balanced or extremely skewed. In our experiments, down-sampling outperformed over-sampling.

---

### #6: Performance measures

We explored popular performance measures to evaluate different systems, i.e. precision, recall, F-measure, and accuracy. We looked at different views on these performance measures of our binary variables

<table>
<thead>
<tr>
<th>Yes Class performance</th>
<th>No class performance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Yes class performance
  - No class performance
  - Average class performance: average of Yes class and No class performance
  - Majority class performance, Minority class performance: if a variable has many Yés, then the Majority class performance is the Yes class performance and vice versa

This step is domain-independent.

---

### Conclusion

In our case study, it is possible to configure the framework to have a system which requires no medical or NLP expertise, and has reasonable results. All the experimented features and methods are general and feasible trapply to different clinical domains. In this study, we showed that the configuration requires:

- Variables (objective function) should be single concept
- Run simple keyword searching first and accept variables whose performance is above a certain threshold.
- For other variables, we try learning machine methods. Re-sampling methods only help minority class classification when the variable is not balanced or extremely skewed.