# Statistical Analysis Plan

## 1 Version

| | |
|---|---|
| TRIAL FULL TITLE | Effects of digitalized differential diagnosis broadening using a computerized diagnostic decision support tool on diagnostic quality in emergency room patients - a multi-centre cluster randomized cross-over trial. |
| SAP VERSION | 1.0 |
| SAP VERSION DATE | 21.03.2023 |
| TRIAL STATISTICIANS | Dr. Thimo Marcin, Prof. Stefan Schauber |
| Protocol Version (SAP associated with) | v.1.5, 10/10/2022 |
| SPONSOR-INVESTIGATOR | Prof. Wolf Hautz |
| SAP AUTHOR | Dr. Thimo Marcin |

## 2 SAP Signatures

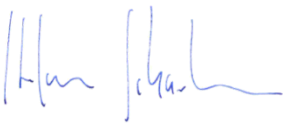
I give my approval for the attached SAP entitled <Effects of digitalized differential diagnosis broadening using a computerized diagnostic decision support tool on diagnostic quality in emergency departments - a multi-centre cluster randomized cross-over trial.>, dated <21.03.2023>

**Statistician (Author)**
Name: Thimo Marcin

Signature: 

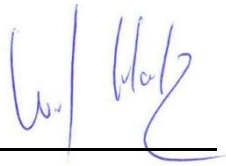
Date: 21.03.2023

**Statistician Reviewer (As applicable)**
Name: Stefan Schauber

Signature: 

Date: 21.03.2023

**Principal Investigator**
Name: Wolf Hautz

Signature: 

Date: 21.03.2023

# 3 Table of Contents

# 4 Abbreviations and Definitions

| | |
|---|---|
| AE | Adverse Event |
| CDDS | Computerized diagnostic decision support systems |
| CRF | Case Report Form |
| DDx | Differential diagnoses |
| ED | Emergency Department |
| GLMM | generalized linear mixed model |
| GP | general practitioner |
| IMP | Investigational Medical Product |
| ITT | Intention to Treat |
| PP | Per Protocol |
| SAP | Statistical Analysis Plan |

# 5 Preface

The purpose of this Statistical Analysis Plan (SAP) is to provide a detailed statement of the intended statistical analyses to be performed on data collected in the DDX-BRO study. The SAP is intended to be a document that stands alone from the study protocol to which it will adhere in the main points of analysis. However, the SAP can undergo revision outside of the protocol. It is not anticipated that revisions of the SAP that are in the spirit of the pre-specified protocol analysis would require review by an ethics committee. The analysis plan also outlines the proposed layout of tables and figures to be presented in the main results paper.

# 6 Study Objectives and Endpoints

## 6.1 Study Objectives

The overall objective of this research project is to reveal the intended and unintended micro-, meso- and macro-level consequences by providing a differential diagnoses (DDx) generator (computerized diagnostic decision support system; CDDS) to physicians in an ED.

The primary (micro-level) objective is to assess the effect of the DDx generator usage on diagnostic quality in patients admitted to the ED. Our primary hypothesis is that ED patients will have a slightly but significant reduced diagnostic quality risk when physicians are asked to consult the provided DDx generator after the first physical examination.

Secondary objectives are

- on the micro-level, to investigate how the use of a DDx generator affects patient related outcomes such as mortality, length of stay, or unscheduled revisits.
- on the meso-level, to understand how DDx generator affect diagnostic workflow in the ED and physicians' advice seeking, collaboration and confidence calibration.
- on the macro-level, to investigate the economic implications of DDx generators utilization in ED and to evaluate the educational implications for physician training.

The safety objective corresponds to the primary outcome, namely how the use of a DDx generator affects diagnostic safety in patients admitted to the ED.

## 6.2 Endpoints

### 6.2.1 Primary endpoint

The primary endpoint is a binary score indicating the risk of reduced diagnostic quality, composed of:

- death within 14d days after ED discharge (yes/no)
- unscheduled medical care (ED revisits, general practitioner (GP) visits or hospitalization) within 14 days after ED discharge (yes/no)
- unexpected intensive care unit admission from ward within 24h when hospitalized (yes/no)
- diagnostic discrepancy between the ED discharge diagnosis and the current diagnosis 14 days after ED admission (yes/no)

The primary endpoint is *true (1)*, if one or multiple of the criteria above become true, and *false (0)* if none of the criteria above occur.

Secondary endpoints include

- all variables of the primary endpoint individually
- unscheduled ED/GP revisits after 72h and 7 days
- length of ED stay
- length of hospital stay if hospitalized
- diagnostic tests in ED
- diagnostic tests after ED
- resource consumption in ED (costs)
- care consumption after ED discharge
- discharge destination
- number and patterns of DDx provided by the physicians
- number of cases where the generated DDx list entails the diagnoses after 14d
- diagnostic error based on full chart review for a random subset
- CDDS usage (timing and number of queries)

Safety endpoints correspond to the endpoint specified in the primary endpoint (except diagnostic discrepancy). Additionally, rates of serious adverse events will be assessed.

Additional outcomes are physician confidence calibration, advice seeking and collaboration assessed by qualitative methods (physician observations, interviews and focus groups) to understand how DDx generator affect diagnostic workflow in the ED and physicians' advice seeking, collaboration and confidence calibration. These outcomes focus namely on physicians and their diagnostic process as a quality assurance. Specification of study endpoints are in chapter 7.5 Study variables.

# 7 Study Methods

## 7.1 General Study Design and Plan

DDX-BRO is an open label, cross sectional, multi-center, six-period cross-over cluster-randomized controlled superiority trial. Participating EDs were randomly allocated to two different sequences of alternating intervention and control periods of two months each. During intervention periods, physicians of the respective ED were asked to query the CDDS for DDx. During the control period, the DDx generator will not be provided to the physician and the diagnostic process will follow the usual care. No wash-out periods will be applied as no substantial cross-over effects are expected given the nature of the intervention.

## 7.2 Inclusion-Exclusion Criteria and General Study Population

Subjects fulfilling all of the following inclusion criteria are eligible for the investigation:

- Informed Consent signed by the subject.
- Presentation to the ED with fever, abdominal pain, syncope or non-specific complaint (NSC) as chief complaint because all these complaints occur frequently, can result from a large number of underlying diseases and thus provide room for diagnostic error. Further, there are no universally agreed algorithms for the diagnostic workup of any of these symptoms (as there is for chest pain, for example). NSC are defined as all chief complaints not included in the check list of specific complaints according to Nemec et al. 2010 (1) or as degraded general condition (german: "reduzierter Allgemeinzustand", "AZ-Reduktion" or similar) as chief complaint.
- Triaged as "no acute life-threatening condition" because study intervention would not be feasible in many cases.
- The study subject is 18 years old or older.

The presence of any one of the following exclusion criteria will lead to exclusion of the participant:

- Trauma as chief complaint, because there are standardized diagnostic workups, most trauma patients receive radiographic imaging and the potential benefit of a DDx generator is questionable.
- Pregnancy (anamnestic), because options for diagnostic workup are severely constrained in these patients and presentation is mostly related to pregnancy and its complications, reducing room for error to occur and be remediated.
- Worsening of a known pre-existing condition or medical referral with a definite diagnosis, because the diagnosis is clear in this case
- Inability to follow the informed consent and investigation procedures, e.g due to language barriers, psychological disorder, admittance via police, detainee status.
- Previous enrolment into the current investigation

## 7.3 Randomization and Blinding

There will be two pre-defined sequences with 2 x 3 alternating intervention and control periods. Participating EDs were randomly assigned to one of the two sequences prior to the recruitment phase, stratified by center size (Insel, Solothurn, vs. Tiefenau, Münsingen). An independent person drew one study site from box 1 (Insel, Solothurn) and one study side from box 2 (Tiefenau, Münsingen) to allocate them to sequence 1 (i.e., Intervention-Control-Intervention-Control-Intervention-Control). The remaining sites will be allocated to sequence 2 (i.e., Control-Intervention-Control-Intervention-Control-Intervention).

Treating physicians cannot be blinded towards the patient's study allocation. However, patients will be blinded, i.e., they will not be informed about the current condition (intervention or control period) of the respective ED site. The study nurses conducting the follow up interviews with patients and their

general practitioners will be blinded and all raters involved in the study will be blinded (when determining whether a diagnostic discrepancy occurred and when conducting chart review to validate the measure of our primary outcome). The statistician performing the primary analysis (Prof. Stefan Schauber) will be blinded. The trial statistician (Dr. Thimo Marcin) will prepare the dataset by replacing the group allocation codes with unspecific codes (group A, group B) before the primary efficacy analyses will be performed by the external statistician.

## 7.4 Study Assessments

| Investigation Periods | Screening | Consent (ICF) | Treatment, Intervention Period | | Follow-up |
|---|---|---|---|---|---|
| Emergency care in the emergency room | Admittance and triage | Waiting time | Medical examination and treatment | | discharged or hospitalized |
| Time | -1 to -5h | -1 to -5h | 0 to 1h | 1h + LOS | 14d±4d |
| **A)** Enrolment | | | | | |
| In- /Exclusion criteria | x | | | | |
| Patient information | | x | | | |
| Patient consent (ICF) | | x | | | |
| **B)** Intervention | | | | | |
| CDDS application during intervention period | | | x | (x) | |
| CDDS usage monitoring by study nurses | | | x | | |
| **C)** **Assessments** (CR: performed within clinical routine as appropriate) | | | | | |
| Demographics | x | | | | |
| Chief complaint | CR | | | | |
| Triage assessment | CR | | | | |
| Medical history | | | CR | | |
| Physical examination | | | CR | CR | |
| Vital signs | | | CR | CR | |
| Laboratory tests | | | CR | CR | |
| Other diagnostic tests | | | CR | CR | |
| CDDS input / output data collection | | | x | (x) | |
| Physician questionnaire | | | x | (x) | |
| Patient telephone interviews | | | | | x |
| Medical record review from ED, hospital and/or GP | | | x | x | x |
| Serious Adverse Events, Adverse device effects | | | x | x | x |
| Device Deficiencies | | | x | x | |
| **D)** **Primary Outcome Score** | | | | | |
| All-cause mortality | | | | | x |
| Unscheduled medical care if discharged (GP, ED revisit, hospitalization) | | | | | x |
| Unexpected ICU admission or upscale in care within 24h if hospitalized | | | | | x |
| Current diagnosis for presenting complaint | | | | x | x |
| **E)** **Secondary Outcomes** | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| Number and cost of ED diagnostic tests | | | | x | x |
| Time to ED diagnosis | | | | x | |
| ED differential diagnoses | | | | x | |
| Physician confidence in ED diagnosis | | | | x | |
| Discharge destination | | | | CR | |
| ED LOS | | | | CR | |
| Hospital LOS if hospitalized | | | | | CR |
| CDDS usage (number of queries) | | | x | (x) | |
| Patient reported outcomes | | | | | x |
| ED, Emergency room; ICU, intensive care unit; CDDS, computerized diagnostic decision support system; GP, general practitioner; LOS, length of stay; CR, clinical routine | | | | | |

## 7.5 Study variables

### 7.5.1 Primary endpoint

| **Variables** | **Description** | **Scale** | **Time Window*** |
|---|---|---|---|
| Diagnostic quality risk | The score is positive and patients at risk for a diagnostic error if one or multiple of the criteria below becomes true: | 0, not at risk<br>1, at risk | 14±4 days |
| Death | Patient died during time of follow-up | 0, no<br>1, yes | 14±4 days |
| Unscheduled medical care | Patient had at least one medical appointment (GP, ED, Hospital, rehabiliation or others) after ED or hospital discharge that was related to the initial ED presentation but not scheduled at the time of ED or hospital discharge. The criteria is fulfilled if the medical appointment was registered in the hospitals' EHR or reported by the patient or GP during follow-up. | 0, no<br>1, yes | 14±4 days |
| Unexpected Intensive Care Unit admission | Patient was first admitted to the ward and then transferred to the intensive care unit (or next upscale in care) within 24 hours. | 0, no<br>1, yes | up to 24 hours after ED discharge |
| Diagnostic discrepancy between ED diagnosis and diagnosis at follow-up | The primary diagnosis has changed during the time of follow-up and is medically different compared to the primary ED diagnosis. The criterion is not fulfilled if the diagnosis at follow-up is verbatim or medical identical with the primary ED discharge diagnosis, just more precise or a complication, i.e. diagnosis at follow-up was not foreseeable at the time of ED diagnosis. The classification follows a pre-defined scheme as used in a previous study (2). The diagnosis at 14d is based on the documentation of the GP or treating Hospital at the time of follow-up. If patients did not report any medical care after ED, we assume that the diagnosis did not change since ED or hospital discharge. | 0, no<br>1, yes | 14±4 days |
| *Follow-up was performed within 14±4 days. All events between baseline and follow-up were documented including time of event (e.g death date or date of re-visits) | | | |

### 7.5.2 Secondary endpoints

| Variables | Description | Scale | Time Window* |
|---|---|---|---|
| Death | see primary endpoint above | 0, no<br>1, yes | 14±4 days |
| Unscheduled medical care | see primary endpoint above | 0, no<br>1, yes | 14±4 days |
| Unscheduled medical care | see primary endpoint above | 0, no<br>1, yes | 7 days |
| Unscheduled medical care | see primary endpoint above | 0, no<br>1, yes | 3 days |
| Unexpected Intensive Care Unit admission | see primary endpoint above | 0, no<br>1, yes | 24h |
| Diagnostic discrepancy between ED diagnosis and diagnosis at follow-up | see primary endpoint above | 0, no<br>1, yes | 14±4 days |
| Length of ED stay | Time difference between ED admittance to ED discharge or ward admission | number of hours (ranging from 0 up to 24 hours) | up to 24 hours after ED discharge |
| Length of hospital stay if hospitalized | Days between ward admission and hospital discharge | number of days ranging from 0 days (discharged at admission day) to 18 days (if hosptalized at end of follow-up) | up to 18 days |
| Diagnostic tests in ED | Data are sought from the EHR to indicate if the the patient recieved one of the following diagnostic tests during the ED stay: laboratory (blood, urine or sputum), MRI, CT, sonography, x-ray, other. Data is collected for each diagnostic type and corresponding subgroups (e.g abdominal CT, head CT etc.) | 0, no<br>1, yes | length of ED stay, up to 24 hours |
| ED resource consumption (costs) | Hospitals administrative cost accounting data will used as a measure of costs that were caused during patient's ED stay. Total costs and subcategories of the cost unit accounting are collected. | numeric in Swiss Francs (CHF) | length of ED stay, up to 24 hours |
| Total resource consumption (costs) | Hospitals administrative cost accounting data will used as a measure of costs that were caused during patient's ED and hospital stay if hospitalized. Total costs and subcategories of the acost unit accounting are collected. | numeric in Swiss Francs (CHF) | length of hospital stay, up to 24 hours |
| Diagnostic tests after ED | Data are sought from the EHR, GP or patient reported during follow-up to indicate if the patient received one of the following diagnostic tests on one of the medical appointments (see care consumption above). laboratory (blood, urine or sputum), MRI, CT, sonography, x-ray, other | 0, no<br>1, yes | 14±4 days |
| Ddischarge destination | Discharge from ED to one of following locations | 0, home<br>1, hospital (intern)<br>2, hospital (extern)<br>3, nursing home<br>4, rehabilitation<br>88, other | length of ED stay, up to 24 hours |

| | | | |
|---|---|---|---|
| Number of DDx provided by the physicians | Number of differential diagnoses provided in the ER discharge letter or EHR documentation. | numeric | length of ED stay, up to 24 hours |
| Number of cases where the generated DDx list entails the primary diagnosis after 14d | The primary diagnosis at end of follow-up was provided at least once of one of the provided DDx-list resulting from a DDx-Gernerator query. | 0, no<br>1, yes | 14±4 days |
| Diagnostic error based on full chart review for a random subset | For a random subset of 50 patients with and 50 patients without positive primary endpoint, a full chart review will be performed to identify diagnostic errors using the Safer Dx Instrument (3) The Instrument includes 13 statements to the presence of a diagnostic error. | Likert Scale from 1 (strongly disagree) to 7 (strongly agree). | |
| Tracked CDDS usage | All queries of the DDx-Generator are tracked and recorded in the study database.<br>- Number of queries per patient and physician<br>- Number of queries per patient and physician<br>- Time difference in hours between ED admittance and query | numeric | length of ED stay, up to 24 hours |
| Self-reported CDDS usage and impact | Physicians questioned with an online survey regarding the use and impact of the generator in the diagnostic process for the individual patient. The questionnaire includes Nine statements regarding diagnosis, confidence, diagnostics, consultations and general helpfulness. | 0, no<br>1, yes | 0-3 days |
| *Follow-up was performed within 14±4 days. All events between baseline and follow-up were documented including time of event (e.g death date or date of re-visits). | | | |

### 7.5.3 Safety endpoints

| **Variables** | **Description** | **Scale** | **Time Window** |
|---|---|---|---|
| Death | see primary endpoint above | 0, no<br>1, yes | 14±4 days |
| Unscheduled medical care | see primary endpoint above | 0, no<br>1, yes | 14±4 days |
| Unexpected Intensive Care Unit admission | see primary endpoint above | 0, no<br>1, yes | 24 hours |
| SAEs | Number and type of serious adverse events:<br>- Results in Death<br>- Life-threatening illness or injury<br>- Results in permanent disability / incapacity<br>- Medical or surgical intervention to prevent life-threatening illness, injury or permanent impairment<br>- Congenital anomaly / birth defect<br>- Results in chronic disease<br>- Requires hospitalisation or prolongation of existing hospitalisation | 0, no<br>1, yes | 14±4 days |

### 7.5.4 Covariates

| Variables | Description | Scale | Time Window |
|---|---|---|---|
| Group allocation | control or intervention phase at time of inclusion | 0, control phase<br>1, intervention phase | baseline |
| Phase | Number of phase at time of inclusion | Numeric, range 1-6 | baseline |
| Age | Age in years | numeric | baseline |
| Sex | Sex | 1, male<br>2, female | baseline |
| Chief complaint | | 0, fever<br>1, abdominal pain<br>2, syncope<br>3, non-specific complaint | |
| Triage level | Triage level | 2-5 | baseline |
| Admission type | Admission to the ED | 0, Self-admittance<br>1, Ambulance<br>2, Air Rescue<br>88, other | baseline |
| Referral type | Type of referral to the ED | 0, self-referral<br>1, GP<br>2, Hospital (intern)<br>3, Hospital (extern)<br>4, Air rescue (primary)<br>5, Ambulance (primary)<br>88, other | baseline |
| Charlson-Comorbidity-Index | Charlson Comorbidity Index (CCI) is Score of 19 comorbid conditions weighted from 1 to 6 | Numeric, range 0-37 | baseline |
| Confidence | Confidence regarding proposed primary diagnosis. (self-reported by the diagnosing physician via online survey) | Likert scale from 1 = not confident (50%) to 5 = confident (100%) | 0-3 days |
| Difficulty | Perceived difficulty of the diagnostic reasoning process. (self-reported by the diagnosing physician via online survey) | Likert scale from 1 = difficult to 5 = easy | 0-3 days |
| Familiarity | Familiarity/experience with the diagnosis (self-reported by the diagnosing physician via online survey) | Likert scale from 1 = not seen before to 5 = familiar | 0-3 days |
| Typicality | Diagnosing physician judgement if the patient's presentation was atypical for the primary diagnosis. | 0, typical<br>1, atypical | 0-3 days |
| Diagnosing physicians | ED physicians (senior and resident) responsible for the treatment during patient's ED stay are recorded | Unique ID | length of ED stay, up to 24 hours |
| Age (physician) | Age in years | numeric | baseline |
| Sex (physician) | Sex of the diagnosing physician | 1, male<br>2, female | baseline |
| Medical experience (physician) | Time of medical experience since obtaining the medical degree (years) | numeric | baseline |
| Emergency experience (physician) | Time of medical experience in emergency medicine (years) | numeric | baseline |
| Medical specialist (physician) | Medical specialist degree | 0, no<br>1, yes | baseline |

# 8 Sample Size

The sample size calculation has been performed for a multi-period cross-over cluster randomized controlled trial using according to Hemming at al. (2020) using the web-tool The Shiny CRT

Calculator.(4) The trial is designed to have a power of 80% to detect a clinically significant between-condition-difference in the primary outcome of 5 percent points on an alpha level of 0.05.
For the primary outcome, we assumed a positive composite score in 12% of the cases in the control condition. Further assumptions were a cross-sectional sampling and exchangeable correlation structure, an intra cluster correlation between 0.01 and 0.05, a coefficient of variation of cluster size of 0.5 and a 10% lost-to-follow up patients.
With 4 periods and 2 clusters (EDs) per sequence and the minimal sample size to detect a clinically significant between-condition-difference in the primary outcome of 5 percent points on an alpha level of 0.05 with a power of 80% equals to (in average) 74 patients per period and cluster and 1'184 patients in total. The sample size calculation was initially performed for 4 periods. However, the trial has been extended to 6 periods due to slow recruitment without modification of the overall required sample size.

# 9 General Analysis Considerations

## 9.1 Timing of Analyses

Once the study has been completed and all data have been entered and cleaned, a review of the data will be conducted and final changes will be made to this SAP. No efficacy analyses will be performed until the final version of this SAP has been approved. Any post-hoc, exploratory analyses not defined in this SAP will be clearly identified in the final report. Any deviations from the planned analyses detailed in this SAP will be documented with reasons in a post-analysis version of the SAP.

## 9.2 Analysis Populations

### 9.2.1 Intention to Treat (ITT) set

Data from all participant with or without protocol violation including dropouts and withdrawals will be included this analysis population.

### 9.2.2 Per Protocol (PP) set

Patients from the intervention group will be removed from the PP-analysis if no CDDS query has been documented. Vice versa, patients from the control group will be removed from the analysis if physicians self-report the query of any DDx generator outside the study protocol

### 9.2.3 Safety set

All subjects who received any study treatment (including control) and completed the study. Patients who did not complete follow-up will be excluded, except reason for non-completion is death.

### 9.2.4 Others:

None

## 9.3 Covariates and Subgroups

Subgroup analyses and treatment interactions will be assessed with following covariates/subgroups:

- Patient characteristics (Age, sex, chief complaint)
- Disease characteristics: confidence, difficulty, familiarity, typicality
- Physician characteristics: medical experience, emergency experience, medical specialist

All analyses on interactions and subgroups are explorative and described in section 11.3 or 13.

## 9.4 Missing Data

### 9.4.1 Missing outcome data

The amount of missing data for the primary endpoint is anticipated to correspond to the number drop-outs, withdrawal and lost-to-follow up (10-15% expected).

### 9.4.2 Missing covariate data

We expect very low missing covariate data as used in primary efficacy analyses. Missing rates of 5% and lower were discussed to have no consequences on the estimates. (5) Therefore, missing data will be handled with full information maximum likelihood estimation. In case of a missing data rate of above 5%, we will repeat the primary efficacy analysis after multiple imputation as sensitivity analysis.

## 9.5 Interim Analyses

An interim analysis for safety outcomes was planned for the end of the second period.

As specified in the clinical investigation protocol (v1.5), safety outcomes are:

- Death within 14±4 days after ED (yes/no)
- Unexpected IMC admission within 24h after ED discharge if hospitalized (yes/no)
- Unscheduled medical care for the same complaints within 14±4 days after ED discharge (yes/no)
- Serious Adverse Events

### 9.5.1 Purpose of Interim Analyses

Based on the evaluation, the sponsor investigator and local PIs decided whether the clinical trial has to be stopped or continued

### 9.5.2 Planned Schedule of Interim Analyses

After study end of the last included patient of period 2.

### 9.5.3 Stopping Rules

No formal stopping rules were set a priori. Decision is based on clinical judgement of the sponsor - investigator and local PIs.

### 9.5.4 Analysis Methods

Safety outcomes were compared between intervention and control periods and between sites using descriptive statistics.

### 9.5.5 Practical Measures to Minimize Bias

Safety interims analysis is performed by one of the trial statistician and data manager (Dr. Thimo Marcin) who is unblinded at any time of the trial. However, the safety interim analysis does not include any efficacy analysis of the primary endpoint and the final analysis will be performed by an external statistician not involved in any study procedure (Prof. Stefan Schauber). Therefore, no additional measures are required to minimize bias (see section 15).

### 9.5.6 Documentation of Interim Analyses

Snapshot of the data available at the safety interim analysis, script for statistical analysis and report for the sponsor and principal investigators is preserved.

## 9.6 Multiple Testing

The primary endpoint is combined score of multiple secondary endpoints, for which we performed power analysis and determine the required sample size to detect a clinically relevant and statistically significant treatment effect. For the single analysis, no multiple testing needs to be performed. However, multiple sensitivity, subgroup and other hypothesis generating analyses will be performed (See section 11 Analysis). We do not plan any adjustments for multiple testing but rather discuss the issue of multiple testing in the limitation of the corresponding papers as appropriate.

# 10 Summary of Study Data

All continuous variables will be summarized using the following descriptive statistics: n (non-missing sample size), mean and standard deviation or median and quartiles as appropriate. The frequency and

percentages (based on the non-missing sample size) of observed levels will be reported for all categorical measures. In general, all data will be listed, sorted by site, treatment and subject. All summary tables will be structured with a column for each treatment in the order (Control, Experimental) and will be annotated with the total population size relevant to that table/treatment, including any missing observations. Summary reporting is subject to change based on Journals' requirement.

## 10.1 Subject Disposition

### 10.1.1 Screening

Number and details (inclusion and exclusion criteria) of screened but not included patients are centrally recorded in a separate database (Research Electronic Data Capture; https://www.project-redcap.org) of the Department of Anaesthesiology and Pain Medicine (KAS) at the Bern University Hospital.

### 10.1.2 Randomization

Sites were randomly allocated to sequences of interventional and control phases before study begin. Subsequently, a calendar with dates defining intervention and control phases for each site was set. Patients were allocated by the study personal to the control or intervention group [cdds_phase] depending on the date they presented to the ED.

### 10.1.3 Non completion

Reasons for non-completion of the study [non_comp_reason] are documented with following categories:

1 Patient withdrew consent
2 Patient was lost to follow-up
3 Patient was excluded requested by PI
4 Patient was excluded due to SAE
5 Patient died
88 Other reason

Patients who died before follow-up will not be considered as lost to follow-up. All others will be considered as lost to follow-up or drop outs.

### 10.1.4 Reporting

The study design is a cluster randomized cross-over trial, however, in- and exclusion criteria were set on subject (i.e patient) and not on group (i.e site) level and subjects included in one period only. Therefore, we deemed a flow diagram on subject level following the CONSORT Flow Diagram for a parallel-randomized trial as appropriate. However, we will additionally present supplemental figures showing inclusion rates over time for the individual clusters.

## 10.2 Derived variables

Several relevant endpoints and covariates are not directly collected in the eCRF but are derived or calculated from source data recorded in the eCRF. The most relevant are shown below, which will be imported in the eCRF before closure of the database

| | |
|---|---|
| Diagnostic quality risk (Primary endpoint) [diagnostic_risk] | True (1) if one of the variables is true (1):<br>- Death [death]<br>- Unscheduled medical care [revisit_yn]<br>- Unexpected Intensive Care Unit admission [icu_admission]<br>- Diagnostic discrepancy between ED diagnosis and diagnosis at follow-up [diagnosis_change_yn] |
| Death [death] | True (1) if patient did not complete the study due to death ([non_comp_reason] = 5) |

| | |
|---|---|
| Unscheduled medical care [revisit_yn] | True (1) if at least one of the documented medical care appointments was unplanned [cc_type_exp] and related or probably related [cc_type_rel] to the initial ED visit. The medical care is classified as "related" if any relation to the initial ED visit at enrollment was indicated by the patient, GP or medical report. |
| Diagnostic discrepancy between ED diagnosis and diagnosis at follow-up<br><br>[diagnosis_change_yn] | True (1) if the diagnoses from the ED discharge and 14d follow-up are 0, identical; 2, precision; 4, complication. False (0) if the diagnoses are 1, diagnostically different; 3, hierarchically different [diagnosis_change] |
| Length of ED stay [los_er] | Difference in hours between [discharge_datetime] and [admission_datetime] |
| Length of hospital stay [los_hospital] | Difference in days between [discharge_hosp_date] and [admission_datetime]. Zero hospital days will be assigned to patients who were not hospitalized directly after ED [discharge_dest] |
| Period [period] | The number of period is derived from the patient's date of signed informed consent [ic_date] and beginning of the respective period as pr-especified for each site.<br><br>Startdate<br>Period 1 2 3 4 5 6<br>Tiefenau 06.06.2022 08.08.2022 10.10.2022 12.12.2022 13.02.2023 17.04.2023<br>Insel 13.06.2022 15.08.2022 17.10.2022 19.12.2022 20.02.2023 24.04.2023<br>Solothurn 20.06.2022 22.08.2022 24.10.2022 26.12.2022 27.02.2023 01.05.2023<br>Muensingen 27.06.2022 29.08.2022 31.10.2022 02.01.2023 06.03.2023 08.05.2023 |
| Charslon-comorbidity index [cci] | The CCI is a score of comorbid conditions that are each weighted from 1 to 6 based on disease severity. Definition of disease severity follows Kim et al. (2014) (6) and derived from following variables:<br><br>[cci_mi], [cci_chf], [cci_pvd], [cci_cva], [cci_dementia], [cci_copd], [cci_ctd], [cci_pud], [cci_hemi], [cci_ckd], [cci_leuk], [cci_lymph], [cci_aids], [cci_liver], [cci_dm], [cci_dm_2], [age] |
| Tracked CDDS usage | Number and timing of queries (relative to ED entry) and number of entered symptoms are derived from [cdds_request_datetime], [cdds_user], [cdds_input_symptoms]. Duplicated queries (no differences in entered symptoms) will not be considered. |

## 10.3 Protocol Deviations

Protocol deviations are considered major if the CDDS were not used during interventional periods or used during control periods. This protocol deviations are addressed by the PP-analysis and sub-analysis described in section **Fehler! Verweisquelle konnte nicht gefunden werden.** and **Fehler! Verweisquelle konnte nicht gefunden werden.**. Follow-up phone interviews may have been conducted after 14±4 days because some patients returned the call after the timeframe or specifically requested a call to on one of the subsequent days. We will perform a primary efficacy sensitivity analysis without these patients if the number of cases exhibits more than 5%. We do not expect any other protocol deviations that may interfere with the primary efficacy analysis.

## 10.4 Demographic and Baseline Variables

All baseline variables as specified in 7.5.4 Covariates

## 10.5 Concurrent Illnesses and Medical Conditions

Diagnoses will be classified according to clinical modification of the International Statistical Classification of Diseases and Related Health Problems (ICD-10 GM)

## 10.6 Treatment Compliance

Number of queries and the average number of symptoms entered into the DDx generator per patient will be used for treatment compliance. In addition, self-reported usage of CDDS outside the study protocol [othercdds1] will be reported.

# 11 Efficacy Analyses

Data will be summarized by treatment groups. N, mean and standard deviation will summarize continuous efficacy variables, whereas number and percent will number and percent will summarize categorical efficacy variables. Statistical analyses will be based on multi-level general linear mixed modelling (GLMM) methods using appropriate post hoc techniques (e.g for subgroup analyses). Regression coefficients for continuous endpoints and odds ratios for binary endpoints and geometric mean ratios for ln-transformed continuous outcomes will be displayed as measure of strength accompanied by their 95% confidence intervals. Viewing plots along with normality testing (e.g. Shapiro-Wilk) will be used in order to check assumptions for the appropriate use of parametric testing approaches. Transformations to normality for variables not fulfilling normality assumptions will be considered (e.g. log, Box-Cox etc.), while nonparametric testing using counterparts of ad-hoc parametric procedures will also be an option as needed (e.g. Kruskal-Wallis instead of one-way ANOVA, the latter being part of the GLM family). R (R Foundation for Statistical Computing, Vienna, Austria) will be used for data analysis. A test-wise 2-sided p-value of less than 0.05 will be considered statistically significant.

## 11.1 Primary Efficacy Analysis

For the primary binary outcome, presence or no presence of a diagnostic quality risk, a generalized linear mixed model (GLMM) with a binomial distribution family and exchangeable correlation structure will be performed. The GLMM takes into account a random intercept for each site, resident and attending physician. In some cases, patients may be treated by more than one resident or attending physician due to changes in shift. The physicians who were initially involved in treatment will be selected as diagnosing physicians. The condition (intervention and control) and the period (period 1 to 6) will be included as fixed factors under the assumption of equality of carry-over effects. Additionally, chief complaint, age, sex and Charlson Comorbidity Index will be added as covariates. The primary efficacy analysis will be performed on the ITT set.

## 11.2 Secondary Efficacy Analyses

### 11.2.1 Secondary Analyses of Primary Efficacy Endpoint

*a) Per Protocol:*

The same model as for the primary efficacy analysis will be performed on the PP set.

*b) Sensitivity analyses:*

Sensitivity analysis will follow the primary efficacy analysis with or without adjustments in model specification.

- GLMM on multiple imputed data (if missing >5%)
- GLMM without baseline covariates
- GLMM with site as fixed factor
- LMM without random effects
- Robust GLMM
- Generalized Estimating Equations to compare against GLMM

*c) Compliance:*

An additional sub-analysis on the PP set will be performed after excluding patients from the intervention group if none of the CDDS query included more than three entered symptoms/clinical features.

### 11.2.2 Analyses of Secondary Endpoints

Secondary endpoints will be analysed using the ITT set.

*a) Death*

The same model as for the primary efficacy analysis will be performed.

*b) Unscheduled medical care*

The same models as for the primary efficacy analysis will be performed.

*c) Unexpected Intensive Care Unit admission*

The same model as for the primary efficacy analysis will be performed.

*d) Diagnostic discrepancy between ED diagnosis and diagnosis at follow-up*

The same model as for the primary efficacy analysis will be performed.

*e) Length of ED stay*

A generalized linear model (GLM) will assess the treatment effect on patients' LOS [hours] in the ED. LOS are count data and hence, models based on normal distribution are not appropriate. (7) Instead, we will apply a Poisson and a negative binomial regression model to select the better performing model based on dispersion and AIC. Following variables will be included in the model as independent variables: treatment, site, period, age, sex, chief complaint, Charlson comorbidity index, admission type. As sensitivity analyses, we will perform a linear regression model after log transforming the dependent variable.

*f) Length of hospital stay*

Analyses will parallel length of ED stay.

*g) Diagnostic tests in ED*

The same model as for the primary efficacy analysis will be performed to assess the treatment effect on likelihood of performed diagnostic test for each the following categories: blood, urine, sputum, MRI, CT, Sonography and x-ray. Subcategories will be analyzed descriptively.

*h) ED resource consumption (ED costs)*

A GLMM will be performed to assess treatment effect on costs. The dependent variable will be natural log transformed to account for skewness. Period and site as well as patients' age, sex, chiefcomplaint, Charlson Comorbidity Index, triage level and admission type will be added as covariates.
A sensitivity sub-analysis without patients from the Bürgerspital Solothurn will be performed as the administrative accounting data may somewhat differ compared to the other hospitals.

*i) Total resource consumption (total costs)*

The analyses for total resource consumptions follows the analyses for ED resource consumption (see h)

*j) Diagnostic tests after ED*

Diagnostic tests during unscheduled medical care after ED/Hospital discharge will be analyzed using descriptive statistics (counts and proportions).

*k) Discharge destination*

Discharge destination will be analyzed using descriptive statistics. Whether the intervention has an effect on inpatient admission or discharge at home is assessed with the analysis on length of hospital stay (see *length of hospital stay* above).

## 11.3 Exploratory Efficacy Analyses

An interaction term between treatment and the fixed effect covariates will be added to the primary efficacy model. Predicted probabilities for diagnostic risk with 95% confidence intervals will be shown for subgroups/covariates in case of any statistically significant interactions ($p<0.05$). Additionally, interactions will also be performed for the analyses on the individual parameters of the composite primary endpoint score.

Additional exploratory analyses for secondary efficacy analyses or other endpoints may be conducted as appropriate. Any additional exploratory analyses will be appropriately labeled and clearly distinguished from planned analyses when results are reported in the clinical study report.

# 12 Safety Analyses

Serious adverse event incidences will be described using counts and proportions for each event category and treatment group (intervention and control). Each subject will only be counted once and any repetitions of serious adverse events will be ignored; the denominator will be the population size of the respective group. If multiple serious adverse event categories are checked for one single event, the event will be assigned to one category only in following order: Death; life-threatening illness or injury; permanent disability / incapacity; medical or surgical intervention to prevent life-threatening illness, injury or permanent impairment; Congenital anomaly / birth defect; Chronic disease; Hospitalization or prolongation of existing hospitalization.

(Serious) Adverse Events will be listed and described if a relation with the clinical investigational protocol or the medical device under investigation cannot be excluded.

## 12.1 Other Safety Measures

The analyses of other safety endpoints as specified in 7.5.3 are described in 11.2.2.

# 13 Other Analyses

*a) Validation of the diagnostic risk score*

Fifty patients with a diagnostic risk (positive primary endpoint) and 50 patients without a diagnostic risk (negative primary endpoint) will be randomly selected among patients from three of four participating sites that share the same electronic health record system (Insel, Tiefenau, Muensingen). If there are less than 50 cases with a positive primary endpoint, the missing cases will be replaced with patients with a negative risk score. A random sampling function of the statistical software R will be used to determine the subset for the full chart review.

For the selected 100 patients, two independent investigators will rate their agreement on a 7-point likert scale for 13 questions regarding presence or absence of a diagnostic error using the revised Safer Dx instrument (3) Ratings between cases with positive and negative risk score will be compared using unpaired two-samples Wilcoxon test. Additionally, logistic regression models will be performed with the primary endpoint composite score as dependent variable and the rating score as independent variable. Additionally, following variables will be added as covariates: site, group allocation, period, site, patient's age, sex, chief complaint and Charlson comorbidity index.

*b) Monitored CDDS usage*

- Descriptive:
  Average number of queries per patient and user and average number of entered symptoms per query and user will be summarized overall, for sites separately and over time using descriptive statistics and illustrating figures.
- Association with physician characteristics:
  Physician characteristics physician characteristics (7.5.4) related to CDDS usage will be assessed with linear models. The average number of queries per patient and average entered symptoms per query will be calculated for each resident physician and used as dependent

variables. Phase, age, sex, medical experience, emergency experience, medical specialist and site will be added as independent predictors.

*c) Self-reported usefulness and impact of CDDS*

- Descriptive:
  Counts and percentage will be used to describe the impact of the CDDS on diagnosis, confidence, diagnostics, consultations and general helpfulness (See 7.5.4).

- Association with physician and patient characteristics:
  Logistic regression models for the binary endpoints described above will be performed. Physician related variables (age, sex, medical experience, emergency experience, medical specialist) and patient related variables (chief complaint, triage level, charlson-comorbidity index and length of ED stay) as well as site will be added as independent predictors. Predictor estimates will be and presented as Odds Ratios alongside 95% confidence intervals using Forest plots. Models will be repeated without physician related variables to explore the potential influence of missing values in physician related covariates.

*d) Number of cases where a generated DDx list entails the primary diagnosis after 14d*

ICD-10 Codes provided form the CDDS will be automatically compared with the ICD-10 Code of the primary diagnosis at follow-up using a batch tool.(8) Contingency tables for patients with and without change in diagnosis (diagnostic discrepancy) will be reported.

# 14 Reporting Conventions

P-values ≥0.001 will be reported to 3 decimal places; p-values less than 0.001 will be reported as "<0.001". Other statistics will be reported as integers or rounded to one or two decimal places as appropriate.

# 15 Quality Assurance of Statistical Programming

A second statistician will review and independently reproduce the primary analyses and summary statistics. The reviewing statistician will have an overview of the entire analyses and will explicitly check the code producing tables as well as any other pieces of code as desired.

# 16 Summary of Changes to the Protocol and/or SAP

## 16.1 Missing Data

Due to simplicity and the expected low rate of missing data for the primary analyses, we aim to handle missing data first with full information maximum likelihood estimation and use multiple imputation only as sensitivity analysis.

# 17 References

White IR and Thompson SG (2005). Adjusting for partially missing baseline measurements in randomized trials. *Statistics in Medicine,* **24,** 993-1007

1. Nemec M, Koller MT, Nickel CH, Maile S, Winterhalder C, Karrer C, u. a. Patients Presenting to the Emergency Department With Non-specific Complaints: The Basel Non-specific Complaints (BANC) Study. Acad Emerg Med. März 2010;17(3):284–92.

2. Hautz SC, Schuler L, Kammer JE, Schauber SK, Ricklin ME, Sauter TC, u. a. Factors predicting a

change in diagnosis in patients hospitalised through the emergency room: a prospective observational study. BMJ Open. 11. Mai 2016;6(5):e011585.

3. Singh H, Khanna A, Spitzmueller C, Meyer AND. Recommendations for using the Revised Safer Dx Instrument to help measure and improve diagnostic safety. Diagnosis. 26. November 2019;6(4):315–23.

4. Hemming K, Kasza J, Hooper R, Forbes A, Taljaard M. A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator. Int J Epidemiol. 1. Juni 2020;49(3):979–95.

5. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res. Februar 1999;8(1):3–15.

6. Kim SM, Kim MJ, Jung HA, Kim K, Kim SJ, Jang JH, u. a. Comparison of the Freiburg and Charlson Comorbidity Indices in Predicting Overall Survival in Elderly Patients with Newly Diagnosed Multiple Myeloma. BioMed Res Int. 2014;2014:1–11.

7. Fernandez GA, Vatcheva KP. A comparison of statistical methods for modeling count data with an application to hospital length of stay. BMC Med Res Methodol. Dezember 2022;22(1):211.

8. Hautz WE, Kündig MM, Tschanz R, Birrenbach T, Schuster A, Bürkle T, u. a. Automated identification of diagnostic labelling errors in medicine. Diagnosis. 11. Mai 2022;9(2):241–9.

# 18 Listing of Tables, Listings and Figures

Here, we provide anticipated tables and figures for the main manuscript. Tables and figures may be subject to minor changes based on Journals' requirement. Additional figures and tables may be included in the manuscript or as supplemental material, based on additional exploratory analyses and Journal requirements.

## 18.1 Figure 1 – Flow diagram

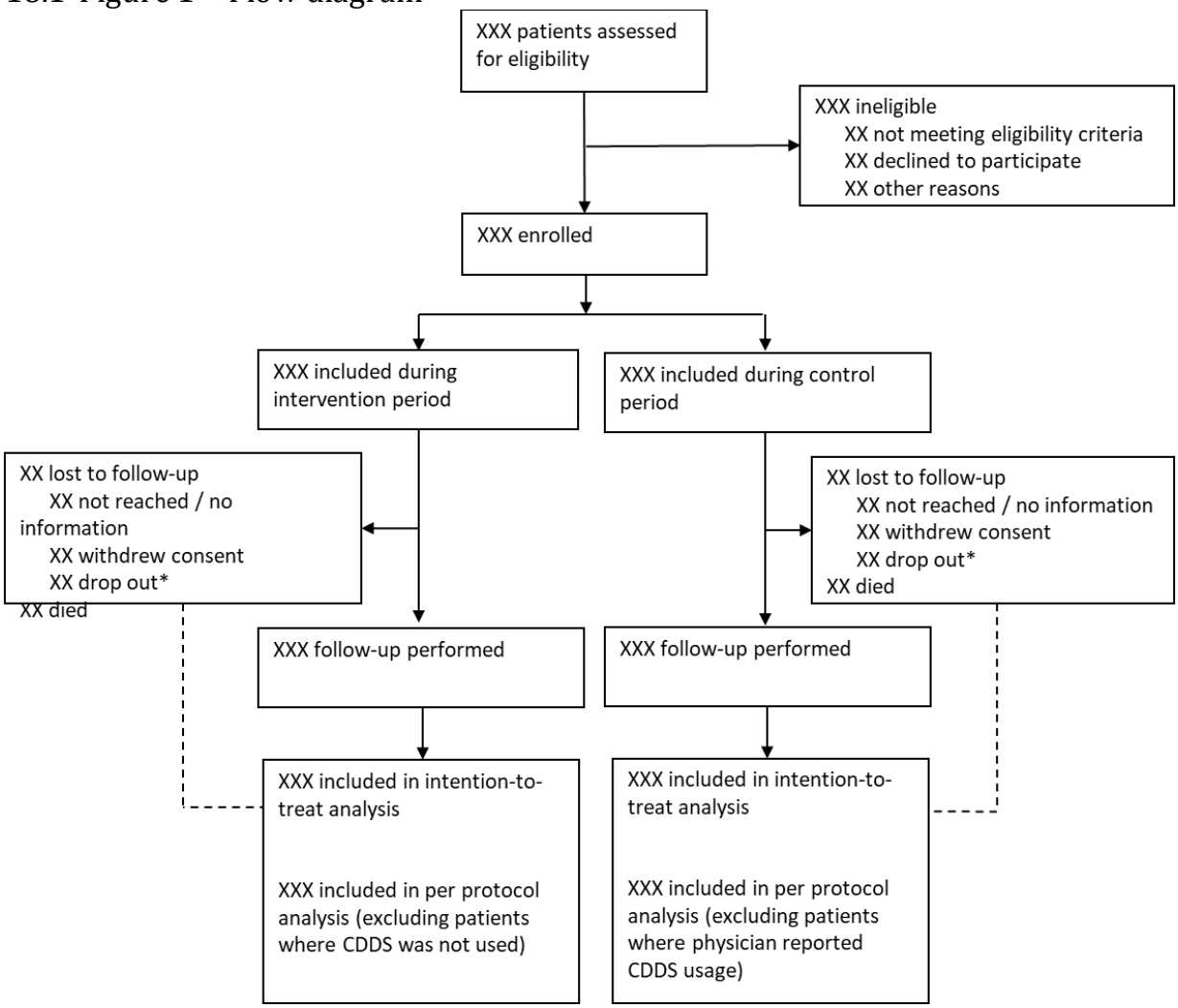



*Figure 1 Trial profile*

** Drop out if patient left ED before clinical examination or if no physician was allocated to the enrolled patient at the end of the study nurse's shift)*

## 18.2 Table 1 – Baseline Characteristics of the intention-to-treat population

| | Intervention Group (n=XXX) | Control Group (n=XXX) |
|---|---|---|
| Site | | |
| University Hospital Bern | XXX (XX%) | XXX (XX%) |
| Buergerspital Solothurn | XXX (XX%) | XXX (XX%) |
| Spital Tiefenau | XXX (XX%) | XXX (XX%) |
| Spital Muensingen | XXX (XX%) | XXX (XX%) |
| Phase | | |
| 1 | XXX (XX%) | XXX (XX%) |
| 2 | XXX (XX%) | XXX (XX%) |
| 3 | XXX (XX%) | XXX (XX%) |
| 4 | XXX (XX%) | XXX (XX%) |
| 5 | XXX (XX%) | XXX (XX%) |
| 6 | XXX (XX%) | XXX (XX%) |
| Age (years) | XX (XX-XX) | XX (XX-XX) |
| Sex | | |
| Male | XXX (XX%) | XXX (XX%) |
| Female | XXX (XX%) | XXX (XX%) |
| Chief complaint | | |
| Fever | XXX (XX%) | XXX (XX%) |
| Abdominal pain | XXX (XX%) | XXX (XX%) |
| Syncope | XXX (XX%) | XXX (XX%) |
| Non-specific complaint | XXX (XX%) | XXX (XX%) |
| Triage level | | |
| 2 | XXX (XX%) | XXX (XX%) |
| 3 | XXX (XX%) | XXX (XX%) |
| 4 | XXX (XX%) | XXX (XX%) |
| 5 | XXX (XX%) | XXX (XX%) |
| Referral type | | |
| Self-referral | XXX (XX%) | XXX (XX%) |
| General practitioner | XXX (XX%) | XXX (XX%) |
| Hospital | XXX (XX%) | XXX (XX%) |
| Air rescue | XXX (XX%) | XXX (XX%) |
| Ambulance | XXX (XX%) | XXX (XX%) |
| Unknown | XXX (XX%) | XXX (XX%) |
| Admission type | | |
| Self-admittance | XXX (XX%) | XXX (XX%) |
| Ambulance | XXX (XX%) | XXX (XX%) |
| Air Rescue | XXX (XX%) | XXX (XX%) |
| Unknown | XXX (XX%) | XXX (XX%) |
| Charlson Comorbidity Index | | |
| 0 | XXX (XX%) | XXX (XX%) |
| 1-2 | XXX (XX%) | XXX (XX%) |
| 3-4 | XXX (XX%) | XXX (XX%) |
| ≥5 | XXX (XX%) | XXX (XX%) |
| Unknown | XXX (XX%) | XXX (XX%) |
| Data are median (IQR) or n (%). | | |

## 18.3 Table 2 – Primary and secondary efficacy outcomes

| | Intervention Group | | Control Group | | Measure of effect | Adjusted effect (95% CI) | p value |
|---|---|---|---|---|---|---|---|
| | N | n (%) or median (IQR) | N | n (%) or median (IQR) | | | |
| **Primary outcome** | | | | | | | |
| Diagnostic quality risk | XXX | XX (XX%) | XXX | XX (XX%) | OR | X·XX (X·XX to X·XX) | <X·XXX |
| | | | | | | | |
| **Secondary outcomes** | | | | | | | |
| Death | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| Unscheduled medical care | XXX | XX (XX%) | XXX | XX (XX%) | OR | X·XX (X·XX to X·XX) | <X·XXX |
| Unscheduled medical care within 7 days | XXX | XX (XX%) | XXX | XX (XX%) | OR | X·XX (X·XX to X·XX) | <X·XXX |
| Unscheduled medical care within 3 days | XXX | XX (XX%) | XXX | XX (XX%) | OR | X·XX (X·XX to X·XX) | <X·XXX |
| Unexpected ICU admission or upscale in care | XXX | XX (XX%) | XXX | XX (XX%) | OR | X·XX (X·XX to X·XX) | <X·XXX |
| Diagnostic discrepancy | XXX | XX (XX%) | XXX | XX (XX%) | OR | X·XX (X·XX to X·XX) | <X·XXX |
| | | | | | | | |
| ED costs in CHF | XXX | XX (XX to XX) | XXX | XX (XX to XX) | GMR | XX (XX to XX) | <X·XXX |
| Total hospital costs in CHF | XXX | XX (XX to XX) | XXX | XX (XX to XX) | GMR | XX (XX to XX) | <X·XXX |
| | | | | | | | |
| **Safety Outcomes** | | | | | | | |
| Serious adverse event (overall) | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| Death | | | | | | | |
| Life-threatening illness or injury | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| Permanent disability / incapacity | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| Medical or surgical intervention to prevent life-threatening illness | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| Injury or permanent impairment | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| Congenital anomaly / birth defect | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| Chronic disease | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| Hospitalization or prolongation of existing hospitalization | XXX | XX (XX%) | XXX | XX (XX%) | | | |
| OR, Odds Ratio; GMR, Geometric Mean Ratio; ICU, intensive care unit; ED, Emergency department, CHF, Swiss Francs | | | | | | | |

## 18.4 Table 3 - Full-case review for diagnostic errors in a random subset. Items of the Safer Dx instrument compared between patients not randomly

| | | **50 random patients with diagnostic quality risk** | **50 random patients without diagnostic quality risk** | **p value** |
|---|---|---|---|---|
| Item 1 | The documented history was suggestive of an alternate diagnosis, which was not considered in the diagnostic process. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 2 | The documented physical exam was suggestive of an alternate diagnosis, which was not considered in the diagnostic process.* | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 3 | Data gathering through history, physical exam, and review of prior documentation (including prior laboratory, radiology, pathology or other results) was incomplete, given the patient's medical history and clinical presentation. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 4 | Alarm symptoms or "Red Flags" (i.e. features in the clinical presentation that are considered to predict serious disease) were not acted upon. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 5 | The diagnostic process was affected by incomplete or incorrect clinical information given to the care team by the patient or their primary caregiver. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 6 | The clinical information (i.e. history, physical exam or diagnostic data) should have prompted additional diagnostic evaluation through tests or consults. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 7 | The diagnostic reasoning was not appropriate, given the patient's medical history and clinical presentation. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 8 | Diagnostic data (laboratory, radiology, pathology or other results) available or documented were misinterpreted in relation to the subsequent final diagnosis. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 9 | There was missed follow-up of available or documented diagnostic data (laboratory, radiology, pathology or other results) in relation to the subsequent final diagnosis. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 10 | The differential diagnosis was not documented OR The documented differential diagnosis did not include the subsequent final diagnosis. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 11 | The final diagnosis was not an evolution of the care team's initial presumed diagnosis (or working diagnosis). | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 12 | The clinical presentation at the initial or subsequent presentation was mostly typical of the final diagnosis. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Item 13 | In conclusion, based on all the above questions, the episode of care under review has a missed opportunity to make a correct and timely diagnosis. | XX (XX to XX) | XX (XX to XX) | <X·XXX |
| Items were rated from 1 (strongly disagree) to 7 (strongly agree) from two independent raters (kappa = x.x)<br>p value are for group comparison using unpaired two-samples Wilcoxon tests | | | | |

## 18.5 Figure 2 – Interaction Plots (example)

Exploratory analyses may be presented using interaction plots.