Student-centred learning with near-peer tutoring compared with a standard faculty-led course for undergraduate training in abdominal ultrasound (the SIGNATURE trial). A multicentre open-label randomized controlled trial

Statistical analysis plan

Table of contents

1 Administrative Information

This trial is registered via https://www.clinicaltrials.gov under the number NCT04114812.

1.1 SAP version

This is SAP version 1 of 17.01.2020.

1.2 Protocoll version

This SAP is based on the SIGNATURE research protocol, version 1 of 17.09.2019

1.3 SAP revisions

There have been no revisions to date.

1.4 Roles and responsibilities (alphabetically)

- Dr Roman Hari, MD, MME, Institute of Primary Health Care (BIHAM), University of Bern, principal investigator, revision of SAP
- Prof. Dr. Michael Harris, MB BS, FRCGP, MMEd, Department for Health, University of Bath, UK and Institute of Primary Health Care (BIHAM), project partner, revision of SAP
- Dr. Andreas Limacher, PhD, CTU Bern, University of Bern, writing of SAP
- Cand. med. Robin Walter, Institute of Primary Health Care (BIHAM), University of Bern, writing of SAP

1.5 Signatures

Robin Walter Writer of SAP	Andreas Limacher Writer of SAP	Michael Harris Senior Statistician

Roman Hari
Principal investigator

2 Introduction

2.1 Background

Ultrasound has become a bedside tool and clinical skill widely used in internal medicine, primary care and other settings. Medical graduates first encounter it in early residency, so Swiss clinicians and programme directors agree that ultrasound training should become integrated into the undergraduate curriculum. Institutions are changing their curricula accordingly, often supplementing traditional teaching with 'near-peer' tutoring through classes held by advanced peers. Near-peer tutoring has been found to be both effective and cost-effective.
Ultrasound education is lagging behind in Switzerland, where we have no consistent ultrasound training for medical students. The most popular course in postgraduate training is a resource-intensive 21-hour basic course for abdominal ultrasound. However, this is expensive, and may not be the best way to impart these skills to undergraduates, who need training more adapted to their needs.
We therefore developed a 21-hour blended-learning ultrasound course, comprising 5 hours of e-learning and 16 hours of near-peer tutoring. Students and their near-peer tutors autonomously organize individual practical teaching sessions within a 16 weeks period. Enrolment started in January 2019. We want to determine whether our curriculum is as good as, or better than, the existing 21-hour course. We also want to better understand how to achieve effective student-centred learning supported by near-peer tutoring.

2.2 Objectives

The aim is to compare a new 21-hour blended-learning ultrasound course (5 hours of e-learning and 16 hours of near-peer tutoring) with the standard 21-hour faculty-led basic ultrasound course.

3 Study methods

3.1 Trial design

The SIGNATURE trial will be a multicentre, open-label randomized controlled non-inferiority trial with two parallel arms, Figure 1.
The students in one arm will have only the new blended learning course, whereas the other arm will have only the standard faculty-led course.

3.2 Randomization

We will use centralized computed randomization in blocks of 4 to allocate each of the participants to one of the two study arms, with allocation in a 1:1 ratio, stratified by study site. The reason for block randomisation is the requirement of the control group to comprise of exactly 24 or 28 students due to the max. group size of 4 students per tutor. The randomization will be performed by an external partner (CTU Bern). An independent statistician of CTU Bern, who is not otherwise involved in the trial, will receive a list of participants of each site that contains the participant ID but no identifying data and will randomly allocate entries to either treatment arm. The generated allocation sequence will be password-protected and sent back to the trial coordinator. This procedure will ensure concealment of

allocation, i.e. randomization will be done without the influence from investigators or study personnel.

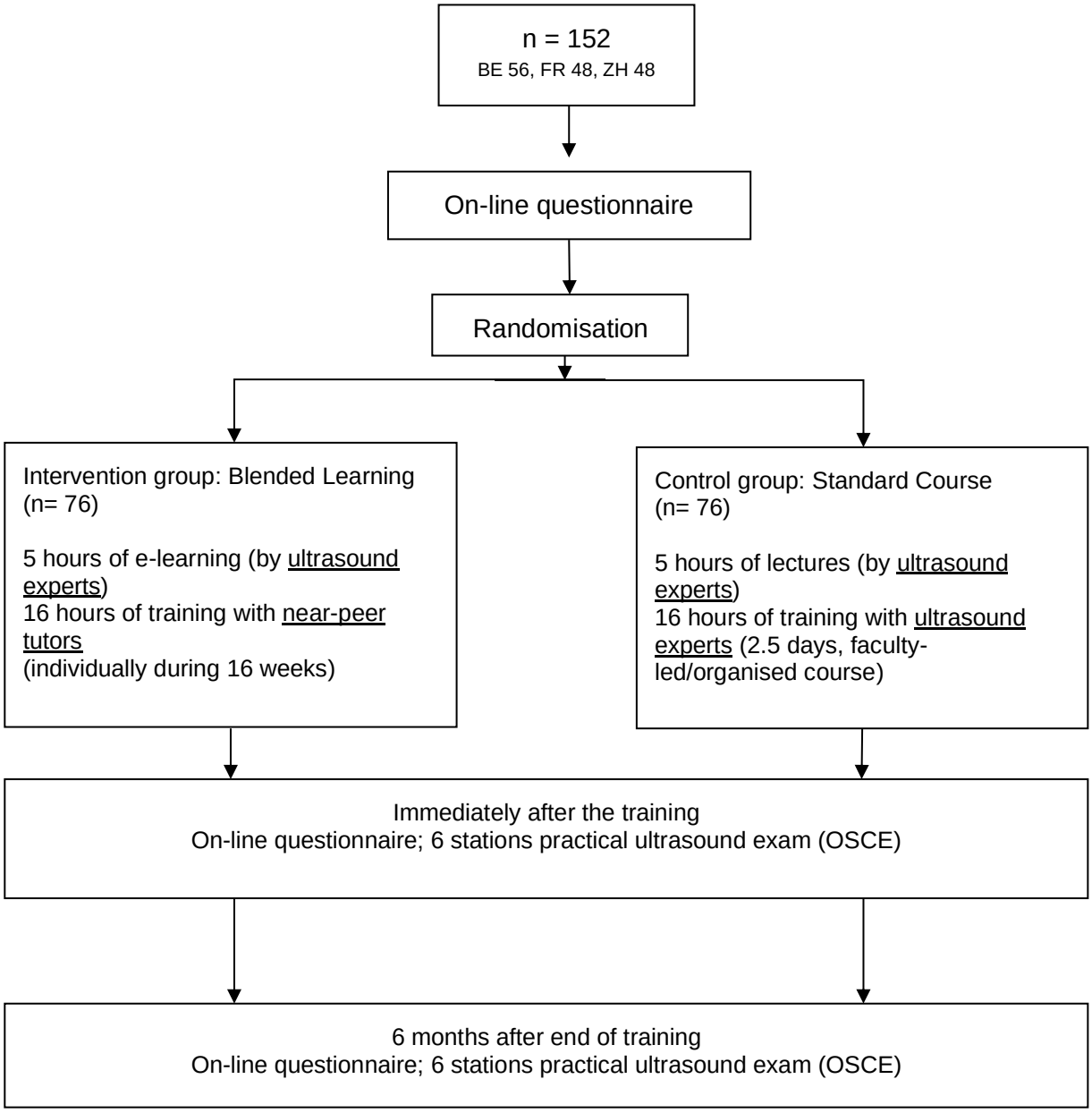

Figure 1: Study methods summary

3.3 Sample size

Based on a pilot with 20 students we expect a mean score of 33 out of 50 points per station with a standard deviation (SD) of 4 points. The study was powered for noninferiority using a margin of -1.5 points for the difference in mean OSCE score between the intervention and control group in the practical exam at 6 months (primary outcome). A difference of 1.5 points would correspond to an effect size of 0.375 standard deviations (SD), which we would still consider as non-inferior. We calculated the sample size with the assumption that the intervention group will

perform somewhat better (+0.5 points in mean OSCE score) than the control group based on test
results from previous pilots. We chose a one-sided alpha-error of 0.05 and a power of 90% which yielded 138 participants (69 in each arm). Allowing for a 9% dropout rate we decided to include 152 students.

3.4 Framework

This is a non-inferiority trial. The working hypothesis is that students from the blended learning group will perform as good or better than the students that took part in a standard basic abdominal ultrasound course 6 months after finishing the course.

3.5 Statistical interim analysis

No statistical interim analyses are planned.

3.6 Timing of outcome assessment

Assessment of the outcomes will be as follows:

- immediately after the course programmes:
    - OSCE scores;
- 6 months after the end of the course programmes:
    - OSCE scores (primary outcome);
    - number of optional post-course training hours taken.

To adapt to local curriculum timetabling, the assessments for the three sites will not be performed at the same time.

3.7 Timing of final analysis

- After completion of the 6-month OSCEs at all three sites.

3.8 Blinding

OSCE assessors will be blinded to group allocation. The assessments will be at the same time for each group. Students will be instructed not to reveal their group allocation to the OSCE assessors.

4 Statistical principles

4.1 Confidence intervals and p-values

Non-inferiority will be assessed based on a one-sided 95%-confidence interval for the difference in the primary outcome, the mean OSCE score at 6 months, between the two groups. If the lower confidence limit lies above the non-inferiority margin of -1.5 points, we will claim non-inferiority.
If non-inferiority can be established, we will test for superiority of the blended learning group at a two-sided alpha level of 0.05. This stepwise testing approach will keep the overall family-wise type-I error rate at the nominal 5%-level.
For secondary outcomes, a two-sided p-value of 0.05 will be considered statistically significant. Results will be reported with corresponding two-sided 95% confidence intervals. There will be no adjustments for multiple testing of secondary outcomes.

4.2 Adherence and protocol deviations
Participants will be considered adherent to the interventions when they have completed the allocated 21 hours of training.
Non-adherence and protocol deviations will be summarized as number and frequency.

4.3 Analysis population

4.3.1 Full analysis set (FAS)
The full analysis set (FAS) will include all randomized subjects. Following the intent-to-treat principle, subjects will be analyzed according to the intervention they are assigned to at randomization.

4.3.2 Per-protocol (PP)
The per-protocol population consists of all subjects in the FAS who do not have any protocol deviation that could confound the interpretation of analyses conducted on the FAS. The following are common major protocol deviations:

- More than 5 hours of previous formal ultrasound training
- Additional practical ultrasound training during the phase between randomisation and first OSCE (other than the training within the study)
- Not completing the assigned training
- Not attending both ultrasound exams

5 Trial population

5.1 Screening data
We will invite all students that belong to a certain study year at the respective university (typically 120-240). The year is selected with regards to best compatibility of the respective study year timetable at a given university with the study procedures. The invitation is sent out during a face-to-face short presentation following a regular lecture of the respective study year as well as through student forums and social media plattforms of the respective study year.
Based on the eligibility criteria, students that decide to take part in the study are expected be more motivated to learn ultrasound than non-participating students. There may also be some students that are not eligible because they are already too advanced in ultrasound skills.
We will not collect any data on non-participating candidates.

5.2 Eligibility
Students of human medicine at one of three Swiss universities (Bern, Freiburg, Zurich, in the 3rd to 8th semester) are eligible for study participation if they are:

- willing to participate in the study and give their consent to take part in the research;
- willing to pay the 200.- CHF course fees (partial reimbursement of tutors and administrative fee for the SGUM course certificate);
- fill out the baseline questionnaire.

Exclusion criteria for study participation are:

- Did not sign the study agreement or pay the course fee;
- More than 5 hours of previous formal ultrasound training;

- Failure to fill out baseline questionnaire.

5.3 Recruitment

We will primarily recruit students in semester 7 in Bern (total medical student intake per year = 220), in semester 6 in Zurich (240) and Fribourg (120).
A CONSORT participant flow diagram will be drawn following the CONSORT 2010 standards (http://www.consort-statement.org/consort-2010).

5.4 Withdrawal and follow up

Participants who do not complete their allocated 21 hours of training will be considered not to have completed the intervention. They will still undergo both OSCEs and will be analyzed on an intention-to-treat-basis, but excluded from the per-protocol analysis.
Participants that did not complete the allocated training or did not attend the two ultrasound examinations (OSCE) will be shown in the CONSORT participant flow diagram.

5.5 Baseline characteristics

We will collect the following baseline characteristics:

- Demographics: age (numeric), gender (F/M), semester (numeric), university (categorical)
- Previous ultrasound experience: owning an ultrasound textbook (yes/no), hours of previous practical experience in ultrasound, hours of formal teaching in ultrasound.
- Preference for future specialisation (list of options plus 'other' option).

Baseline characteristics will be descriptively summarized showing number and proportion for categorical data and mean and standard deviation or median and interquartile range for continuous data. No p-values will be displayed because any significant difference can be explained by the play of chance if the randomization was performed properly.

Table 1: Baseline table.

Description	Variable	Type
Age	age	Numeric/continuous: years
Gender	gender	Categorical/binary: Male, Female
Semester	Semester	Numeric/discrete, 1-12
Universtity	University	Categorical/nominal: Bern, Freiburg or Zürich
Formal ultrasound training	Formal training	Numeric/continuous, hours
Previous ultrasound experience	Experience	Numeric/continuous, hours (Watching, doing)

6 Analysis

6.1 Outcome definitions

6.1.1 Primary outcome

The primary outcome will be students' ultrasound skills measured by a six-station OSCE six months after the end of their courses (mean score).
To assess the outcome, an objective structured clinical examination (OSCE) has been adapted from the validated version of Hofer et al., 2012. Every station in the six-station OSCE has a maximum of 50 points.

6.1.2 Secondary outcomes

- students' OSCE scores immediately after their courses (mean score); see description above
- additional hours of self-training during the follow-up phase;

Table 2: Derivation of primary and secondary outcomes.

Outcome	eCRF sheet	Variable	Variable type	Derivation	Outcome type
Primary: OSCE score six months after course	Excel sheet output from electronic data entry (iPad)	Subscores at OSCE stations 1-12	Numeric	Sum of scores at individual stations (0-50 pts) divided by number of stations	Continuous
OSCE score immediately after course	Excel sheet output from electronic data entry (iPad)	Subscores at OSCE stations 1-12	Numeric	Sum of scores at individual stations (0-50 pts) divided by number of stations	Continuous
Additional hours of self-training	Online Survey before 2nd OSCE (RedCap)	Survey questions asking on additional self-training	Numeric	Simple number between e.g. 0-100 hours	Continuous

6.2 Analysis methods

6.2.1 Primary analysis

As recommended by the CONSORT statement for non-inferiority trials, we will perform both intention-to-treat (ITT) and per-protocol (PP) analyses. For the primary outcome, ITT and PP analyses must reach the same conclusion to establish non-inferiority. For secondary outcomes, the ITT analysis will be the primary analysis, the PP analysis a secondary analysis.
We will use a repeated-measures mixed-effects linear model with robust standard errogrs to compare the mean OSCE scores of both groups 6 months after the course (primary outcome) as well as immediately after the course (secondary outcome), taking into account the possible dependence of data within sites, participants, and series. The model will include the averaged OSCE score over all six stations of a series of each participant at both time points as dependent variable, fixed terms for the learning group, the type of series, the time point and the interaction term between learning group and time point, and random terms for site and participant. We will assess non-inferiority based on a one-sided 95% confidence interval for the difference in mean OSCE score between the two groups 6 months after the course. If the lower confidence limit lies above the non-inferiority margin of -1.5 points, we will claim non-inferiority. If non-inferiority can be established, we will test for superiority of the blended learning group at a two-sided alpha level of 0.05. This stepwise testing approach will keep the overall family-wise type-I error rate at the nominal 5%-level.

The difference in mean OSCE score immediately after the course will be assessed from the same model described above and presented with a two-sided 95% confidence interval and corresponding p-value.
The difference in median number of additional hours of self-training will be assessed from a quantile regression using robust standard errors to account for dependence of data within sites, and presented with a two-sided 95% confidence interval and corresponding p-value.

6.2.2 Secondary analyses

We will perform an additional PP analysis for secondary outcomes.

6.2.3 Sensitivity analyses

As a sensitivity analysis, we will analyse the two time points separately using the same model as described above but omitting fixed terms for the time point and the interaction between time point and learning group as well as the random term for participant. In case the primary model should not fit, we will use this approach as the primary analysis.
Morover, we will analyze individual OSCE scores of each station instead of the averaged OSCE scores over all six stations within a series. We will use the same model as described above, but using a random term for station instead of a fixed term for series. We will also consider random terms for series and assessor as well as a random slope if there is non-zero variability and if the model fits. Should models not fit, we will analyse the two time points separately.
Depending on the distribution of the additional hours of self-training, we will furthermore use a non-parametric Wilcoxon-rank-sum test, a mixed-effects negative binomial model considering zero-inflation and a random term for sites, or a negative binomial model considering zero-inflation and robust standard errors for this secondary outcome.

6.2.4 Additional analyses

We will explore the association of different factors such as gender, hours of additional training, motivation for ultrasound and mean group size with OSCE scores six months after the course using multivariable mixed-effects linear models with robust standard errors.

6.2.5 Assessment of statistical assumptions

We will assess normality and homoskedasticity of residuals of the linear model used to analyse OSCE scores based on a QQ-plot as well as a fitted-value vs. residual plot. If the distribution severely deviates from a normal distribution, we will perform adequate transformation of OSCE scores.

6.3 Missing data

If the OSCE score of at least one station at either time point (immediately after or six months after the course) is available, repeated-measures models properly account for missing data. If more than 5% of students have completely missing

outcome data at both time points, we will perform multiple imputation in the primary analysis, and perform an available case analysis as sensitivity analysis. Chained equations will be used to impute missing data, using predictive mean matching for continuous/ordinal variables, logistic regression for binary variables, and multinomial logistic regression for categorical variables. The following variables will be considered in the imputation model: OSCE score, site, station, hours of additional training, motivation for ultrasound, age, gender, and other baseline variables listed in Table 1. The two learning groups will be imputed separately. We will generate 50 imputed data sets, which will be analysed and combined using Rubin's rule.
For the secondary outcome of additional hours of self-training, we will also use multiple imputation if the primary outcome was analysed using multiple imputation. Else, we will disregard missing data and perform an available case analysis.

6.4 Harms

While no harm to participants is anticipated, if there is a pathological finding in a student a reporting system will be set up to coordinate the control of the finding and the further management. For any reported trial data, these findings will be anonymised.

6.5 Statistical software

Analyses will be performed with Stata version 16 or later (StataCorp, College Station, TX, USA).

6.6 Subprojects

We plan the following subprojects adjacent to shis study

- A study on the effect of (readiness for) self-directed learning
- A study on confidence calibration and its development over time (CROCUS)
- A study on participant satisfaction and perceived advantages and disadvantages of the two educational strategies

The respective research questions and statistical analyses will not be further described in this statistical analysis plan.

6.7 Quality control

A second statistician will perform a quality check of the analysis scripts and statistical report.

6.8 References

6.8.1 Publications

Hofer, M., Kamper, L., Miese, F., Kropil, P., Naujoks, C., Handschel, J., & Heussen, N. (2012). Quality indicators for the development and didactics of ultrasound courses in continuing medical education. *Ultraschall Med, 33*(1), 68-75. doi:10.1055/s-0031-1281649