Preoperative airway assessment – Experience gained from a multicentre cluster randomised trial and the Danish Anaesthesia Database

Anders Kehlet Nørskov

This review has been accepted as a thesis together with four original previously published papers by University of Copenhagen 12th of Marts 2015 and defended on 17th of April 2015

Tutors: Lars Hyldborg Lundstrøm, Charlotte Vallentin Rosenstock and Jørn Wetterslev.

Official opponents: Anders Perner, Annmarie Touborg Lassen and Niklas Nielsen.

Correspondence: Department of Anaesthesiology and Intensive Care Medicine Nordsjællands Hospital, Dyrehavevej 29, 3400 Hillerød, Denmark

E-mail: anderskehlet@hotmail.com

Dan Med J 2016;63(5)B5241

The thesis is affiliated to the Department of Anaesthesia and Intensive Care Medicine, Copenhagen University Hospital, Nordsjællands Hospital, Denmark and
Copenhagen Trial Unit, Copenhagen University Hospital, Rigshospitalet, Denmark.

THE PH.D THESIS IS BASED ON THE FOLLOWING FOUR ORIGINAL PAPERS

Paper I
Nørskov AK, Rosenstock CV, Wetterslev J, Astrup G, Afshari A, Lundstrøm LH. Diagnostic accuracy of anaesthesiologists' prediction of difficult airway management in daily clinical practice: a cohort study of 188 064 patients registered in the Danish Anaesthesia Database. Anaesthesia 2015, 70;272-281.

Paper II
Nørskov AK, Rosenstock CV, Wetterslev J, Lundstrøm LH. Incidence of unanticipated difficult airway using an objective airway score versus a standard clinical airway assessment: the DIFFICAIR trial – trial protocol for a cluster randomized clinical trial. Trials 2013, 14:347.

Paper III
Nørskov AK, Lundstrøm LH, Rosenstock CV, Wetterslev J. Detailed statistical analysis plan for the difficult airway management (DIFFICAIR) trial. Trials 2014, 15:173.

Paper IV
Nørskov AK, Wetterslev J, Rosenstock CV, Afshari A, Astrup G, Jakobsen JC, the DIFFICAIR trial author group, Lundstrøm LH. Effects of using the Simplified Airway Risk Index versus usual airway assessment on unanticipated difficult tracheal intubation – a cluster randomised trial with 64,273 participants. British Journal of Anaesthesia 2016, Accepted for publication

BACKGROUND

'The most compelling educational effort for the anaesthesia community should be to reduce the frequency and severity of complications related to managing the airway'
Jonathan Benumof 1995

The difficult airway

Optimal oxygenation and ventilation of the anaesthetised patient is a core service for the anaesthesiologist. Undergoing general anaesthesia, the patient is commonly deprived of spontaneous breathing following the induction of potent anaesthetic drugs. Hereafter, the anaesthetist re-establishes sufficient ventilation and oxygenation. Thus, a period of apnoea occurs while the provider takes over the breathing. The most commonly applied methods of oxygenation is ventilation through a tracheal tube, a laryngeal mask, or a face mask [1, 2]. Usually, establishment of sufficient ventilation is uncomplicated, reducing the period of apnoea to a minimum, which is easily tolerated by the patient. Difficulties with airway management place the patient at risk of a prolonged period of apnoea and thus, at increased risk of airway related morbidity and mortality. Deprivation of oxygen may result in serious adverse events such as anoxic brain damage, heart ischemia, heart failure and ultimately death [3–5]. However, airway management difficulties may also cause minor adverse events such as tooth injury or vocal cord injury, e.g. due to multiple attempts of instrumenting the airway [6].
The aforementioned methods of airway management may serve as each other's escape strategies, thus oxygen may still be provided to the patient if one or even two methods fail. Nevertheless, it takes time to acknowledge failed ventilation and subsequently change method of airway management, hence increasing the period of apnoea and the risk of adverse events. The incidence of failed intubation is approximately 1 in 1,000 and the incidence of cannot intubate, cannot ventilate is approximately 1 in 2,800-20,000 [1, 7]. The incidence of failed laryngeal mask placement is above 1% and may be even more frequent [8]. Impossible mask ventilation is reported in approximately 1 in every

690 patients [7]. Depending on the definition, 2 to 8% of all intubations turn out to be difficult [9–11]. The incidence of difficult mask ventilation is approximately 0.5-1.5%, and there is a clear correlation between difficult intubation and difficult mask ventilation and the combination occurs in approximately 1 in every 250 patients [12, 13].
Thankfully, these cases are rare and general anaesthesia is a safe and trusted procedure. Nevertheless, when things go wrong the consequences can be catastrophic and with millions of patients undergoing general anaesthesia every month around the globe, this topic – rightly – draws a lot of attention.

Difficult intubation and difficult mask ventilation
There has been many proposals of definitions of difficult intubation and difficult mask ventilation [7, 14–19]. Unfortunately, no internationally accepted definitions exist and several studies still employ the laryngeal view proposed by Cormack and Lehane as a surrogate for difficult intubation [20]. Through the last decade, definitions of difficult intubation and difficult mask ventilation have been programmed into the DAD. The definition of difficult intubation is in keeping with the Canadian Airway Focus Group and the definition of mask ventilation difficulty is based on the definition proposed by Han and colleagues [15, 21]. Throughout the papers comprised in this thesis we have employed the same definitions of difficult intubation and difficult mask ventilation as programmed in the DAD. A change of intubation equipment or more than 2 intubation attempts was regarded as a difficult intubation. Difficult mask ventilation was defined as impossible, inadequate, unstable or requiring two providers
(Figure 4).

Preoperative airway assessment
Prediction of difficult airway management remains a pivotal challenge in anaesthesia and it is highly prioritized among anaesthesia personnel to identify patients at risk of airway management difficulties. Unanticipated airway difficulties may cause a stressful situation in an environment where sufficiently competent personnel and equipment may not be readily available. Correct prediction of the difficult airway alters the potentially dangerous unanticipated airway to an anticipated difficult airway with, predominantly, ample time for proper preparation. Thus, accurate prediction of difficult airway management may reduce potential complications by the allocation of experienced personnel and by using relevant equipment and well planned strategies [22].
In the UK in the late eighties and early nineties the National Confidential Enquiry into Patient Outcome and Death did several reports on perioperative deaths and pointed out the importance of a preoperative assessment and identification of patients at risk of airway difficulties [23]. It has been internationally accepted, that the preoperative assessment should include a thorough assessment of the patient's airway and a subsequent risk assessment of potential airway management problems. All major anaesthesia societies, as well as the Danish, recommend preoperative airway assessment [15, 17, 24]. However, a large British survey published in 2011 (The National Audit Project 4 (NAP4)) found 133 cases of airway related death or severe complications (e.g. brain damage) throughout the UK over a one year period in 2008/2009. Only 35 (26%) of these cases had a formal preoperative airway assessment recorded [24]. One of the main recommendations from the NAP4 was to perform thorough preoperative airway assessment on all patients and to have a plan A, B and C ready for airway management, before instigating anaesthesia. Though increased attention over the last decades, and a general agreement about the need and rationale for a preoperative airway assessment, it is still unclear as to how this assessment should be performed. The American Society of Anesthesiologists (ASA) recommends a preoperative airway assessment based on eleven anatomical variables [17, 25]. However, they do not elaborate regarding, which factors are mandatory for examination, nor on how they should be weighted in an overall airway assessment. The ASA argues that the decision to assess some, or all risk factors depends on the clinical context [17]. Consequently, it is left to the discretion of the individual anaesthesiologist. Likewise, the UK based NAP4 gives no elaboration on the content of airway assessment [2].
Several papers have sought to identify and develop valid tools for prediction of airway management difficulties. Traditionally, the diagnostic accuracy of a predictive test is denoted by sensitivity and specificity. High sensitivity and specificity would indicate a good predictive test. But, an inherent challenge can arise when trying to predict rarely occurring events (e.g. difficult intubation ≈ 5%). Despite developing a test with high sensitivity and specificity, a relatively high number of false positives may be encountered (since the condition is rare), thereby reducing the positive predictive value of the test [26]. Several predictive tests for difficult intubation has demonstrated a positive predictive value at approximately 25-40%, meaning that even amongst the patients expected to be difficult to intubate, the majority will not pose difficulties [27–30]. However, if we were to regard anticipation of intubation difficulties as a 'disease' with an effective 'treatment' (e.g. change of intubation modus) the number needed to treat (NNT) would be 3-4 patients in order to avoid one (unanticipated) difficult intubation [31]. It can be argued, that this is an acceptable number. But then another concern arises: Is it discomforting, stressful or resource requiring to be (wrongfully) categorised as expected difficult to intubate? It might be all of the above. Nonetheless, it can be argued that the discomfort and resources related to, e.g. enhanced focus on positioning; pre-oxygenation; use of advanced intubation equipment; and allocation of experienced personnel may be negligible compared to the benefits of avoiding a potentially life-threatening situation. Thus, the acceptable ratio of true/false positives always has to be considered in the context of the severity of the condition (the harm/benefit ratio). The positive likelihood ratio is an alternative statistic for assessing diagnostic accuracy and it is defined as the sensitivity/(1-specificity). It estimates how much the odds of an event (e.g. difficult intubation) increase in case of a positive test (e.g. anticipation of difficult intubation) [32]. If the positive likelihood ratio of a test is high (generally above 10) the test may be relevant to perform, even though its sensitivity may just be moderate.
Despite acknowledging the value of current preoperative airway assessment tests, it may be possible to further improve the predictive value of airway assessment in general, thereby further reducing the number needed to treat. The NAP4 recommends uniformities on airway assessment. It seems reasonable to assume that implementation of a rigorous and systematic airway assessment approach for all patients undergoing anaesthesia would be superior to usual standards of care. It would require a large multicentre trial to compare rigorous and routine use of the best available standards for airway assessment with usual care [1, 31, 33]. Further, it would involve a firm infrastructure and widespread dedication from the providers [31]. The firm infrastructure is present in Denmark, as the Danish Anaesthesia Database may

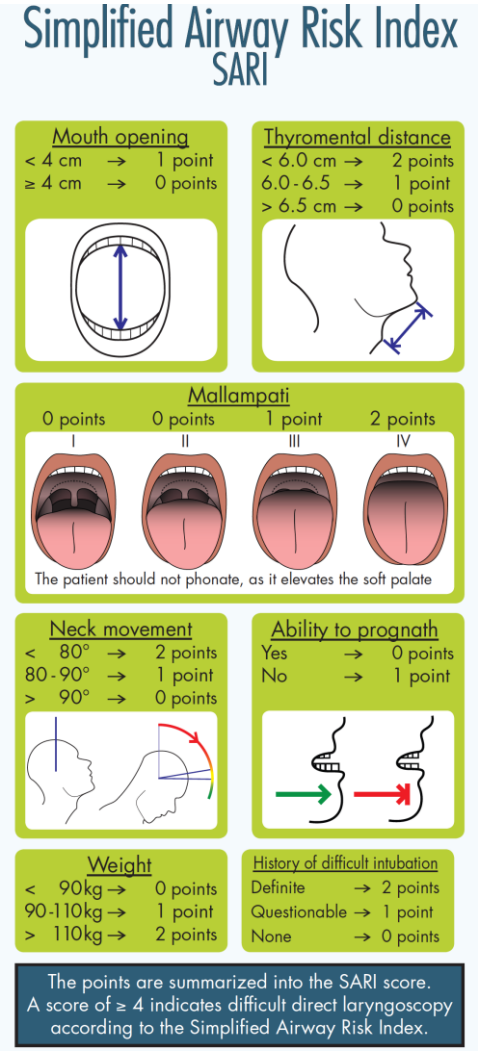
serve as the registration platform and the Central Civil Registry (CPR) enables unique identification of individual patients.

Predictive models

No single predictor is sufficiently valid in predicting difficult intubation or difficult airway management in general [9, 11, 34–36]. However, several studies indicate that by combining multiple predictors of difficult intubation the predictive value of the assessment increase [11]. Many multivariable risk models for prediction of difficult intubation have been proposed [27–30], yet none have been developed using state of the art methodology. Therefore, they contain potential risks of systematic error (bias) and random error i.e. type 1 and 2 errors [33, 37].

As it is often the case with risk– and prognostic models they have not been sufficiently tested in a relevant clinical setting versus usual care on the field. [33, 38, 39]. The 'Simplified Airway Risk Index' (SARI) is a multivariable model for airway assessment described by El-Ganzouri and colleagues [27] (Figure 1). It enables an estimation of the likelihood of a difficult direct laryngoscopy.

Figure 1. The (modified) Simplified Airway Risk Index used in Paper 4



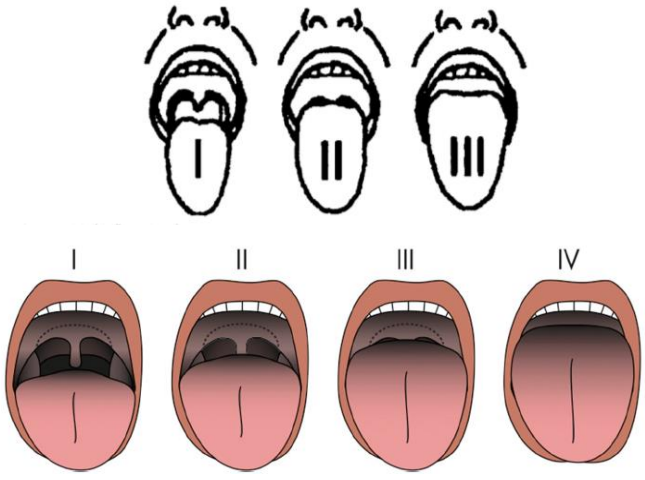
The SARI consists of 7 independent risk factors for difficult intubation. (1) Mouth opening, (2) thyromental distance, (3) Mallampati grade, (4) neck movement, (5) ability to prognath, (6) weight and (7) history of difficult intubation. Each risk factor is assigned a weighted score of 0-1 or 0-2 points. A summarised score (the SARI score) of ≥ 4 is indicative of difficult direct laryngoscopy in the original publication. In Denmark a modified Mallampati comprising four classes has been widely accepted, whereas the original Simplified Airway Risk Index was developed using the original Mallampati grading, ranging from 1-3 [40, 41] (Figure 2).

We decided to adhere to the known procedure in Denmark and including the modified Mallampati classification in a slightly modified SARI model in Paper 4. When the SARI was constructed (as often opted when constructing predictive models), the authors chose to dichotomise or otherwise categorise continuous variables, leading to potential loss of information; an increased risk of false positives; and concealment of any potentially non-linear relation between variables and outcome [42]. The SARI model was developed from a large population, albeit never externally validated nor internally validated using bootstrapping methods. This induce risk of overestimating the predictive value of the model [38]. Though the methodology used in developing the SARI was not flawless, we found the SARI to be the best suitable available model, to test in a clinical trial. The SARI has several important strengths: it was developed from a large study material; it is quick to perform; and easily learned and implemented in a clinical setting.

Preoperative airway assessment in Denmark before initiation of this PhD

The present PhD study was commenced in 2011 and, as in the USA and the UK, there was no clear recommendation on how to perform airway assessment in Denmark [43]. Consequently, we assumed that in daily routine practice, prediction of difficult intubation was based on the individual anaesthesiologist's response to the following question: do I anticipate a difficult intubation? [13]. The answer to that question may or may not be based on a diverse array of preoperative airway examinations, depending on the individual anaesthesiologist and departmental recommendations.

Figure 2. The original Mallampati grade (top) and the modified classification (buttom)



Prior to engaging in a trial testing a predictive model it was important to establish whether preoperative airway assessment was being performed in Denmark, and if there was any kind of uniformity. We investigated departmental recommendations on airway assessment and found a wide discrepancy between Danish regions and between departments of anaesthesiology [43]. Also, we found that preoperative airway assessment, in some form, was widely practiced in Denmark, and we therefore found it fair to assume, that the anaesthesiologists preoperative assessments were based on one, or several known risk factors of difficult airway management [13] (Figure 3).

Furthermore, there is reason to believe, that there may be a certain variance in the performance of the preoperative assessment from patient to patient, and between physicians within departments. Ultimately, we concluded that prediction of difficult intubation was based on the anaesthesiologist's individual response to the question: do I anticipate a difficult intubation? Previous studies have focused on the predictive value of a single risk factor or the value of combining several risk factors into a multivariable predictive model. The diagnostic accuracy of the individual anaesthesiologists' prediction of airway management difficulties, pragmatically reflecting daily clinical practice, has never been investigated.

The Danish Anaesthesia Database

The Danish Anaesthesia Database is a national clinical quality assurance database containing selected quantifiable indicators, covering the anaesthetic process from the preoperative assessment, through anaesthesia and surgery, until discharge from the postanaesthesia care unit. Most variables, and all airway related variables, are mandatory for registration. Anaesthesiologists have to tick Yes/No boxes to answer two mandatory questions regarding the anticipation of difficult intubation and difficult mask ventilation, following preoperative airway assessment. Additionally, the scheduled airway management plan is recorded. Immediately following airway management, an intubation score is registered based on the actual conditions of the tracheal intubation. An analogue score for mask ventilation is registered for patients who undergo attempts of mask ventilation (Figure 4). Information regarding gender, age, ASA classification, height and weight is also mandatory for registration. Furthermore, the DAD contains information on choice of anaesthesia technique and certain use of drugs.

Figure 3. The number of risk factors printed on anaesthesia records in departments in Denmark in 2012

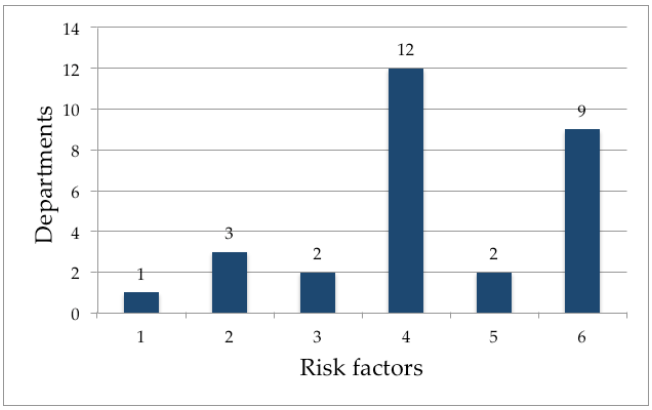
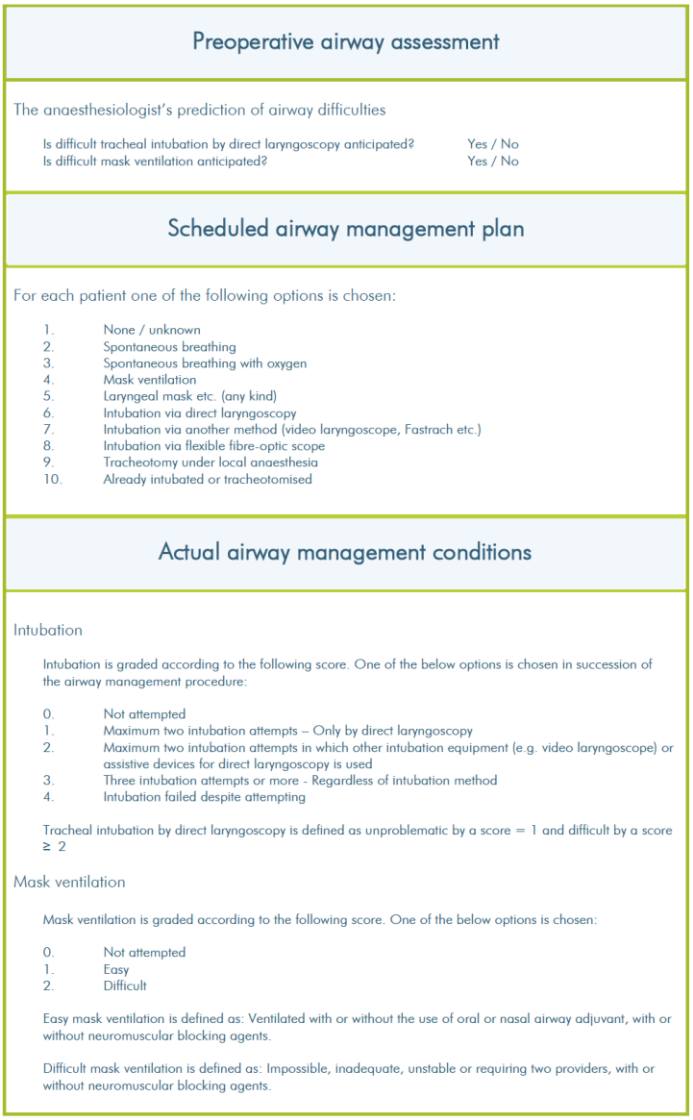


Figure 4. Data registration in the Danish Anaesthesia Database in 2011



Each patient is uniquely registered into the database using a personal identifying number from the Danish Civil Register. The identifying number enables easy identification of each patient, thus reducing the risk of duplicates and other wrong samplings. Using the civil registration number further enables identification of patients anaesthetised multiple times during a defined period. In 2011, 37 departments of anaesthesia recorded data to the Danish Anaesthesia Database. More have joined over the last years and the DAD now covers approximately 80 percent of all patients undergoing surgery and anaesthesia in Denmark. All variables are predefined and links to user manuals are integrated in the DAD interface for each variable. The airway related variables are not registered in other registries and they are therefore not possible to formally cross validate. However, the registration platform comprises several validation and completion rules, securing data completeness and preventing obscure data registration. Most departments use the data for quality assurance and for registration of their productivity, thus reinforcing follow-up registrations on missing patients. Prior studies have proven the DAD to have good patient coverage compared to data from the National Patient Register [44, 45].

Objectives

The overall objective of this PhD thesis was to reduce the incidence of unanticipated difficult airway management. Difficult airway management remains the number one reason for anaesthesia related serious adverse events and the unanticipated difficult airway is associated with an increased risk of morbidity and mortality. It is therefore believed to represent a surrogate for airway related morbidity and mortality.

In order to achieve the objective of reducing the incidence of unanticipated difficult airway management the following part aims were defined:

- To determine the proportion of unanticipated difficult intubation and unanticipated difficult mask ventilation.
- To explore and quantify 'usual care' for preoperative airway assessment in every day practice.
- To investigate and design a trial, based on state of the art methodology for testing a predictive model in a randomized setting.
- To test the implementation of a systematic risk model for preoperative airway assessment versus usual care on the incidence of unanticipated difficult intubation.

PRESENTATION OF PAPERS

PAPER 1

"Diagnostic accuracy of anaesthesiologists' prediction of difficult airway management in daily clinical practice: a cohort study of 188,064 patients registered in the Danish Anaesthesia Database"

Background

All major airway societies recommend preoperative airway assessment [1, 15, 17], yet the choice of content is ultimately at the discretion of the individual anaesthesiologist. The predictive accuracy of this assessment has, to our knowledge, never been evaluated, and the aim of the study was to do so.

Methods

In the Danish Anaesthesia Database we identified 188,064 patients who had undergone tracheal intubation from June 1st 2008 to June 1st 2011 (Figure 5). The anaesthesiologists' preoperative predictions on intubation and mask ventilation difficulties were compared with actual airway management conditions.

Results

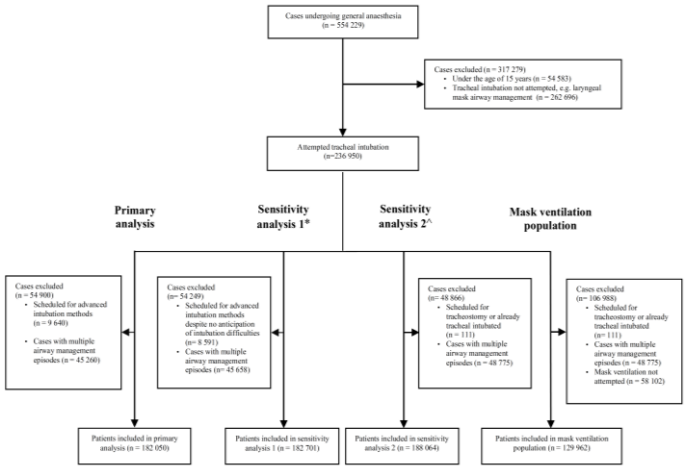
We found a total of 3,383 (1.86%) difficult tracheal intubations, of which 3,154 (93%) were unanticipated. When difficult intubation was anticipated, 229 of 929 (25%) had an actual difficult intubation (positive predictive value). As a consequence of including patients who were anticipated difficult to intubate, scheduled for and intubated using advanced techniques, as true positives in a sensitivity analysis 1, the proportion of difficult intubations that were unanticipated, reduced to 75%.

Difficult mask ventilation occurred in 857 patients (0.66%). It was unanticipated in 808 cases (94%).

Conclusion

The proportion of unanticipated difficult tracheal intubation in daily routine practice, ranging from 75 to 93%, underlines the existing challenges in predicting airway management difficulties. There may be ample room for improvement based on a rigorous, evidence based and systematic approach.

Figure 5. Flow diagram of the study populations of Paper 1



* First sensitivity analysis: This population includes, in addition to the population in the primary analysis, a group of patients predefined as correctly identified difficult intubations. These were patients anticipated difficult to intubate, scheduled for and intubated by more advanced methods than direct laryngoscopy (e.g. video laryngoscopy)

^ Second sensitivity analysis: This population includes all tracheal intubated patients. Difficult intubation is defined as an intubation score ≥ 3.

PAPER 2

"Incidence of unanticipated difficult airway using an objective airway score versus a standard clinical airway assessment: the DIFFICAIR trial – trial protocol for a cluster randomized clinical trial"

Background

Choice and content of preoperative airway assessment in Denmark is ultimately at the discretion of the individual anaesthesiologist. Systematic, evidence-based and consistent airway assessment may reduce the incidence of unanticipated difficult airway management. The Simplified Airway Risk Index [27] is a multivariable risk score for predicting difficult intubation.

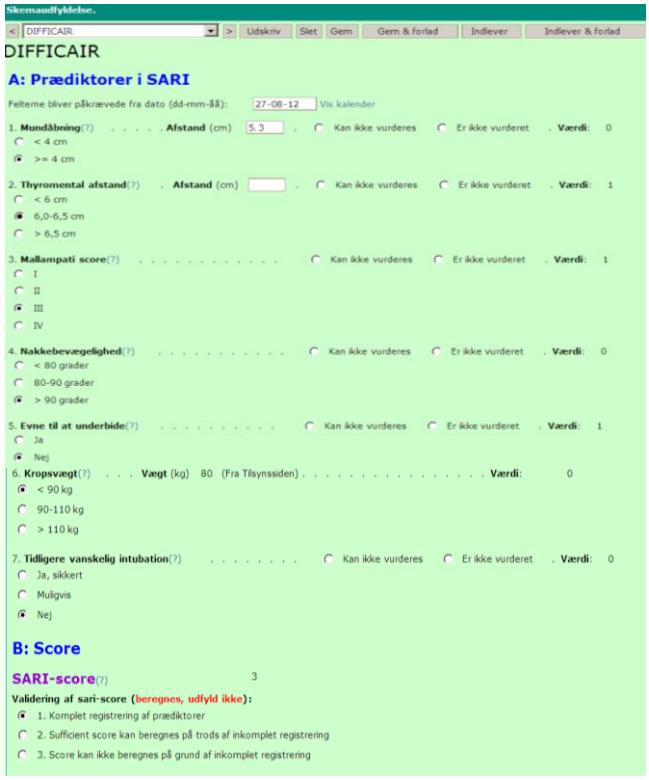
Objective

To compare the effect of implementing the SARI as a systematic airway assessment tool with usual standards for airway assessment, on the incidence of unanticipated difficult intubation. We hypothesised a relative risk reduction of 30%, corresponding to a number needed to treat of 180 patients.

Methods

We 1:1 cluster randomised 28 Danish departments of anaesthesia to airway assessment by the SARI or by usual standards of airway assessment. The primary outcomes were the proportion of participants with unanticipated difficult and unanticipated easy intubation. Main secondary outcomes were 48 hours- and 30 days mortality and statistics for addressing diagnostic accuracy of a test (e.g. sensitivity and specificity). The intervention was a systematic education in use of–, and DAD registration of the Simplified Airway Risk Index (Figure 6 and 7). The usual care departments continued preoperative airway assessment and registration in the DAD as before. To fully address the complexity of the clinical question, we found it necessary to define two different populations. Population 1: patients that were attempted intubated, but not preoperatively scheduled for advanced intubation techniques. Population 2: patients from Population 1 plus patients anticipated

Figure 6. Screen dump (in Danish) of the DIFFICAIR registration interface in DAD on SARI departments



difficult to intubate, scheduled for, and attempted intubated by an advanced technique. Outcomes were assessed for both populations.

Conclusion

In order to enhance transparency of the DIFFICAIR trial the protocol (Paper 2) was made public on www.clinicaltrials.gov (NCT01718561) prior to trial initiation and published in TRIALs. The protocol was written according to the SPIRIT 2013 statement [46].

PAPER 3

"Detailed statistical analysis plan for the difficult airway management (DIFFICAIR) trial"

Background

To prevent outcome reporting bias and data-driven analyses, it is encouraged to prospectively publish a trial protocol [46, 47]. The same argument applies for a prospective publication of a statistical analysis plan.

Method

The statistical analysis plan was written, published on www.clinicaltrials.gov (NCT01718561), and submitted for publication before the last data entry of the DIFFICAIR trial and before any outcome data were extracted.

Figure 7. Preoperative registration in the DAD

Preoperative airway assessment

- Usual care departments -

A: The anesthesiologist's prediction of airway difficulties

Is intubation by direct laryngoscopy anticipated to be difficult? Yes / No

- SARI departments -

A: Predictors in the Simplified Airway Risk Index

1. Mouth opening:

< 4 cm → 1 point
≥ 4 cm → 0 point

In patients with incisors the distance between the teeth is measured at maximum mouth opening. In edentulous patients the intergingivale distance is measured at maximum mouth opening. The distance is measured and recorded in centimeters.

2. Thyromental distance:

< 6 cm → 2 points
6.0 - 6.5 cm → 1 point
> 6.5 cm → 0 points

Measured along a straight line from the "Prominentia Laryngea of cartilago Thyroidea" to the notch of "Mentum Mandibulae" with maximum head extension. The distance is measured and recorded in centimetres.

3. Modified Mallampati class:

I → 0 points
II → 0 points
III → 1 point
IV → 2 points

The visibility of the oropharyngeal structures are assessed on the patient sitting in neutral position with maximum mouth opening and tongue protrusion without phonation.

Class I: Soft palate, fauces, uvula and faucial pillars visible
Class II: Soft palate, fauces and uvula visible
Class III: Soft palate and base of uvula visible
Class IV: Soft palate not visible

4. Neck movement:

< 80 ° → 2 points
80-90 ° → 1 point
> 90 ° → 0 points

The range of motion from full extension through full flexion is categorized as < 80°, 80°- 90° or > 90°. The range is assessed by asking the patient to do a full extension of the neck. Then, the anesthetist places and fixates, a specially designed card in the patient's temporal region in a way that the long side of the card aligns a vertical line e.g. in a window frame. The position of the card in relation to the head is held fixed while the patient does a maximum neck flexion. Subsequently, the position of the card is compared with a horizontal line in the room, for example the window frame.

5. Ability to prognath:

Yes → 0 points
No → 1 point

The capacity to bring the lower incisors in front of the upper incisors. Edentulous patients is categorized as Yes.

6. Body weight:

< 90 kg → 0 points
90 - 110 kg → 1 point
> 110 kg → 2 points

Based on medical records or the patient's own information the weight in kg is recorded.

7. History of difficult intubation:

Definite → 2 points
Questionable → 1 point
None → 0 points

B: The SARI Score

The summarized SARI score was calculated in the Danish Anaesthesia Database

C: The anesthesiologist's prediction of airway difficulties

Is intubation by direct laryngoscopy anticipated to be difficult? Yes / No

General analysis principles

All main analyses will compare the two trial groups using intention-to-treat (ITT) and, in order to ensure a correct type 1 error risk, all main analyses will account for the clustered design of the trial and the stratification variable [48–51]. Sensitivity analyses will be performed adjusted and unadjusted for potential predefined confounding covariates and on predefined populations.

Statistical analysis
The primary analyses of the primary outcomes will be adjusted for the stratification– and the cluster variable in a generalised estimating equation (GEE). The robustness of the results is tested by repeating the analyses in a mixed effects model and with a standard t-test comparing the means of the outcome at department level between trial groups.

Conclusion
We intended to increase the transparency and the robustness of the data analyses by an a priori publication of a statistical analysis plan.

PAPER 4

"Incidence of unanticipated difficult intubation using the Simplified Airway Risk Index versus usual airway assessment – a cluster randomized clinical trial in 64,273 patients – The DIFFICAIR trial"

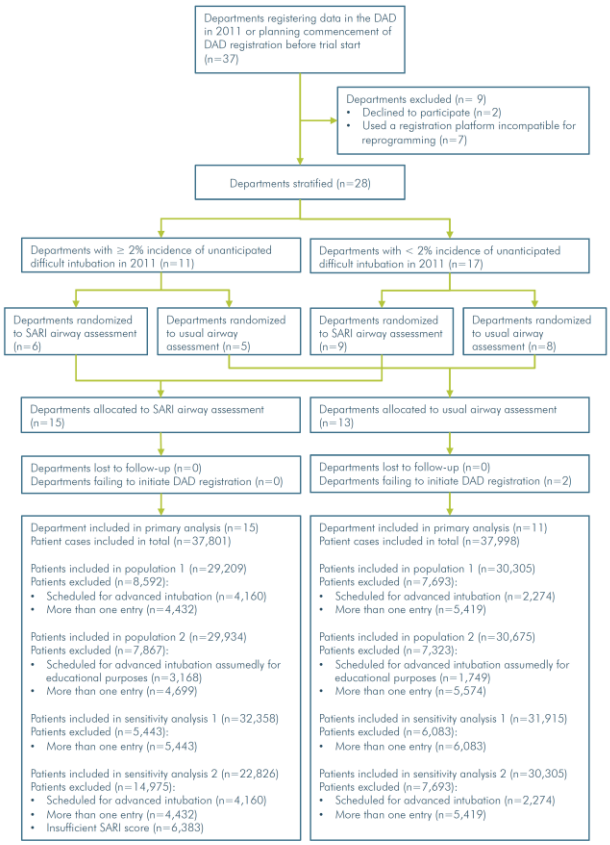
Results
A total of 26 clusters were included (15 SARI departments and 11 usual care departments (Figure 8)).

Primary outcomes
In population 1, 59,514 patients, SARI (29,209) and usual care (30,305), were included. In SARI departments 2.38% (696) of the patients had an unanticipated difficult intubation versus 2.39% (723) in usual care departments. Odds ratio (OR) adjusted for cluster and stratum was 1.03 (95% CI, 0.77–1.38), P = 0.84. The proportion of unanticipated easy intubation was 1.42% (415) in SARI versus 1.00% (302) in usual care departments. Adjusted OR was 1.26 (0.68–2.34), P = 0.47.

Secondary outcomes
We found no statistical significant differences between the trial groups in adjusted secondary outcomes. The SARI departments had a 58% unadjusted increase in patients anticipated difficult to intubate (4.32% versus 2.73%) and an 84% unadjusted increase in patients scheduled for advanced intubation techniques (10.33% versus 5.62%). Adjusted odds ratios did not reach statistical significance.

Conclusion
Applying the SARI compared to usual airway assessment for prediction of difficult intubation did not result in a statistically significant change in the incidence of unanticipated difficult or easy intubation. However, using the SARI may increase the anticipation of intubation difficulties and may change practice towards using advanced intubation techniques.

Figure 8. Flow diagram of the DIFFICAIR trial



DISCUSSION
Paper 1 presents a novel and previously unpublished estimate of the diagnostic accuracy of prediction of difficult airway management in daily clinical practice. The primary outcome was the proportion of difficult intubations being unanticipated. As a consequence of including a group of patients predefined as true positives, the proportion of unanticipated difficult intubation was reduced from 93% in the primary analysis to 75% in the first sensitivity analysis. The 'true' accuracy of the anaesthesiologists' predictions of intubation difficulties is probably somewhere between the predictive values found in the primary analysis and the sensitivity analysis. The primary analysis may have a tendency to underestimate the predictive accuracy and could be regarded as a 'worst case' scenario, whereas the sensitivity analysis may be regarded as a 'best case' scenario, tending to overestimate the predictive accuracy. We found a similar high proportion of 94% of unanticipated difficult mask ventilation, and Paper 1 further underlines the clear association between difficult mask ventilation and difficult intubation reported in previous studies [7, 10]. Furthermore, the proportion of combined difficult mask ventilation and difficult intubation is in perfect alignment with prior findings [1, 10].

We assumed that the anaesthesiologists' predictions were based on one or several known predictors of difficult intubation. However, the diagnostic accuracy of the anaesthesiologists' predictions was poor compared to studies on stand-alone tests and multivariable risk scores [11, 27, 35]. While prior studies have

been conducted under rigorous settings Paper 1, in contrast, reflects everyday clinical practice. Comparisons should therefore be made with caution. The findings underline the importance of always being prepared for unanticipated airway management difficulties in daily clinical practice, and that prediction of difficulties remains a challenging task. We therefore speculated that there might be room for improvement, based on a rigorous, evidence-based and systematic approach.

The DIFFICAIR trial

In Paper 2 we described the innovative use of a national clinical database as the basis for a randomised clinical trial (RCT). In order to present the optimal transparency, the protocol was written according to the SPIRIT 2013 recommendations and published prior to trial commencement [46]. We presented 'state of the art' study methodology for testing the implementation of a multifacetted recommendation [33, 52, 53].

To avoid potential outcome reporting bias and data-driven results paper 3 presents a detailed statistical analysis plan for the intubation part of the DIFFICAIR trial. In order to eliminate falsely low type 1 error rates due to the trial design, our primary outcome analyses were adjusted for the design variables, such as clustering and stratification [54]. We choose to compare the intervention effect using a generalised estimating equation with an exchangeable correlation matrix, in order to account for intra cluster correlation [55, 56].

The value of a diagnostic test is usually presented as sensitivity and specificity. We found it clinically more relevant to present the diagnostic accuracy of the test as 1 - total accuracy. Hence, focusing on the proportion of unanticipated difficult intubations (false negative) and unanticipated easy intubations (false positive). Patients being 'false negative' are at increased risk of hypoxia, airway related morbidity and even death. Although less severe, the 'false positive' patients are at risk of being imposed unnecessary stress and discomfort by for example awake intubation. Furthermore, both categories take up a large amount of potentially unnecessary resources. Since both sensitivity and specificity may be difficult to interpret intuitively, we chose to present more transparent primary outcomes. Additionally, using the proportion of unanticipated difficult intubation allowed us to perform a baseline cohort study, on which we based our sample size estimation in due consideration of the between and within cluster variation of the primary outcome before initiation of the trial. By prespecifying our methods and analyses, we hope that the results from the DIFFICAIR trial will be as transparent and robust as possible.

The intervention in the DIFFICAIR trial was a combination of systematic application of the SARI for all patients, thorough education of physicians and nurses, and mandatory registration of the SARI variables in the DAD. This intervention did not lead to a significant reduction in the proportion of unanticipated difficult or easy intubation. Albeit not reaching adjusted statistical significance, the anaesthesiologists' behaviour tended to change on SARI departments towards an increase in the number of patients predicted difficult to intubate and an increase in the number of patients scheduled for advanced intubation techniques.

General discussion

Throughout this thesis the proportions of difficult intubation and difficult mask ventilation were relatively low compared to previous studies [11, 12, 34, 57]. Since there is no international consensus on how to define difficult intubation or difficult mask ventilation these variables are hard to directly compare between studies [16, 26]. Furthermore, difficult intubation is often equated with, and described by the laryngoscopic view classified by Cormack and Lehane, which is merely a surrogate for difficult intubation [20]. The definitions of difficult intubation and difficult mask ventilation predefined in the DAD have been employed consistently throughout this thesis. To test the robustness of our results, sensitivity analyses using a more rigorous definition of difficult intubation were also performed. Likewise, we found it necessary to perform additional sensitivity analyses on different patient populations in order to fully disclose the complexity of the data.

In Paper 1, the proportion of difficult intubation in the population was 1.86% and in the DIFFICAIR trial the proportion of difficult intubation was 2.66% and 2.62%, respectively (SARI and usual care). A major concern of the DIFFICAIR trial was to enhance focus on correct registration in the DAD. Furthermore, minor revisions were made to the DAD prior to initiation of the DIFFICAIR trial, making the registration of airway related variables easier and more reliable (see below). The increased frequency of difficult intubations from Paper 1 to 4 is most likely attributed to successfully enhancing the quality of the data in the database, rather than an actual increase in difficulties. The increased frequency of events enhanced the power in Paper 4 compared to the sample size estimation based on 2011 data.

In Paper 1, only 47.5% of the patients were scheduled for advanced intubation techniques when intubation difficulties were expected. This number increased substantially in the DIFFICAIR trial to 58.2% in usual care departments and 65.6% in SARI departments. The main aim of predicting a difficult intubation is to avoid airway related morbidity, ranging from simple tooth injury to anoxic brain damage or even death. Increased allocation of patients to advanced intubation techniques may require more resources, e.g. more personnel and use of costly equipment, however it was undoubtedly the right approach for some patients. It is debatable to which ratio the patient related benefits outweigh the harms, especially when harm includes potential major adverse events [31, 58]. Some of these adverse event measures were not accessible in the database, and we cannot rule out that the systematic use of the SARI may have had a beneficial (or harmful) impact on other outcomes when the ones recorded.

As in Paper 1, the predictive accuracies found in the DIFFICAIR trial are not readily comparable with previous study findings. The original SARI was developed from an observational study material and tested on the same material, thus never prospectively validated. The DIFFICAIR trial on the other hand is a randomised trial, testing the implementation of a multi faceted recommendation, affecting every day clinical practice. In alignment with the original SARI publication, prior observational studies on risk factors or risk models for difficult intubation have demonstrated moderate to good predictive accuracy of the examined models. However, they have been conducted under rigorous settings and some even validated on the same population. This induces a substantial risk of exaggerating the prognostic value and an element of publication bias may also exist. Comparison on prediction rates from this cluster randomised trial with prior observational studies should therefore be made with caution.

The Danish Anaesthesia Database

This thesis is based on data from the Danish Anaesthesia Database from 2008 through 2013. The DAD is the largest clinical quality insurance database in Denmark. Its coverage and volume have provided the basis for several observational studies, including Paper 1, and its scale enables research on rare outcomes as difficult airway management. The solid implementation throughout Danish anaesthesia departments made it feasible to use the DAD as the registration platform for the 'case report forms' in a multicentre randomised trial setting (Paper 4). However, no research – observational or randomised – is better than the quality of the recorded data. Thus, a lot of effort has been put into the task of heightening the data quality of the database. In 2011, the database included a few inexpediencies regarding the registration of airway variables, e.g. unfortunate default settings and potential delays for registration of a difficult airway. Therefore, a minor revision was undertaken in conjunction with the programming of the new registration page for the DIFFICAIR trial, and the help interface was updated. Furthermore, a large educational effort was conducted on enhancing focus on correct DAD registration, comprising email distributed tutorials and personal education.

As prior mentioned the definition of difficult intubation is not internationally uniformed and consequently the same applies for the definition of unanticipated difficult intubation. The database does not contain data on the preparations made before intubation, such as having a more advanced intubation device available, and/or having a specialist in anaesthesiology present. Furthermore, the difficult airway is a continuum from minor difficulties to the worst imaginable scenario, the 'cannot intubate, cannot ventilate' situation. The DAD simply allows a dichotomised answer of 'Yes' or 'No' to the questions of anticipation of intubation– and mask ventilation difficulties. Additionally, the intubation score is categorised in the DAD, and in the outcome measures dichotomised, inducing potential loss of information [59, 60]. It would have been preferable to have had more differentiated information on the anticipated and actual difficulties. However, being a clinical tool, the database inevitably has a pragmatic limitation to the extend of data being recorded. When encountering difficult airway management, it is mandatory to fill out the difficult airway management details. Personal vanity, or the reluctance of further registration, may have created an incentive in some personnel to register airway difficulties as less severe than they actually were.

Strengths and limitations

Study 1 was an observational study on patients prospectively entered in the DAD. The study was conducted on a large cohort, reflecting daily clinical practice throughout Denmark from a widespread population of surgical patients, and with a broad span of seniority among anaesthetists. This minimizes the risk of selection bias and increases the external validity, allowing the results to be interpreted in a 'real life' clinical context. Over the 3-year period the proportions of airway difficulties were very stable, reflecting consistent registration practice throughout the study period. The data registrants were unaware of the study being conducted, thus having no direct connection to the investigator group.

By applying a prospective and randomised design, the result of the DIFFICAIR trial would gain a higher level of evidence than results from observational cohort studies [61] (Figure 9).

The DIFFICAIR trial has a number of strengths: 1) Application of state of the art methodology for testing the clinical impact of a predictive model [33, 37], 2) prospective planning and reporting of the trial methodology in a published protocol and statistical analysis plan [62, 63], 3) the applied methodology reduced the risk of systematic error (bias) [64], 4) the risk of random error were limited by including a large number of patients [65], 5) adequate statistical methods were used to account for the clustered nature of the data (GEE) and the robustness of the results was tested in multiple statistical models and through sensitivity analyses and 6) a post hoc analysis on the primary outcome in 2011 data found a perfect baseline balance between the two trial groups.

The main limitation of study 1 is inherent in its observational nature. No certain indication exists for the incentive to allocate patients to a particular airway management technique. For example, it could be for educational reasons; more convenient/less time consuming for the physician; due to tradition; due to lack of other relevant equipment; or because the anaesthetist predicts difficulties with airway management with a certain device. Intuitively, there should be an association between anticipating a difficult airway and scheduling the patient for advanced intubation techniques; allocating experienced personnel to the airway management; or even striving to avoid general anaesthesia. Hence, the indication itself can alter the outcome, e.g. making an otherwise difficult intubation easy, or perhaps instigating another way of handling the airway, not involving intubation of the patient. Moreover, when no difficulties are expected, an otherwise easy intubation may turn out to be difficult, e.g. if least experienced intubator is assigned to the job.

These considerations also apply for the patients in the DIFFICAIR trial. However, the aim of good randomization is random and even distribution of confounders between groups, and when using stratification, evenly distribution of confounders within strata. Furthermore, an effort was made to adjust for any preassumed confounding in the best suitable statistical models and baseline data revealed good pre-trial balance between groups on the primary outcome. Nevertheless, presence of some form of residual confounding can never be entirely ruled out.

Since no other registry records these data, it was not possible to externally validate the airway related data registered in the DAD. Thus, potential unrecognised registration errors are possible. Most departments monitor the registration of patients and do follow-up registrations on missing patients. But, we cannot rule out that some patients, who underwent anaesthesia, were never registered in the database, potentially resulting in an unknown number of missing patients. Since the outcome assessors could not be blinded the person doing the preoperative assessment could potentially also perform the 'assessment' of actual difficulties.

In Paper 4, the enhanced level of education and attention on airway difficulties may have led to an increased awareness and registration of difficult intubations in SARI departments, potentially muddling an effect of the intervention. It was impossible to conduct the trial unnoticed in Denmark and a change of behaviour towards airway assessment resembling the SARI might have happened as a spill over effect on usual care departments. Further, there is a minor risk of contamination bias from the SARI to the usual care departments, e.g. if anaesthesiologists changed work place. Some patients were impossible to assess with the SARI and some anaesthesiologists undoubtedly either forgot or deliberately avoided the use of the SARI. Moreover, we could not ethically dictate the anaesthesiologists to abide by the predictions

of the SARI score. These matters may in some way have obscured a true intervention effect of the SARI.

The risks of error

The reliability of evidence-based medical research is influenced by the risk of three generally accepted levels of error: systematic error ('bias'); random error ('play of chance'); and design error ('wrong design to answer the right question') [66, 67]. Even though we have sought to minimize all levels of error in the DIFFICAIR trial, some dimensions of risk could not be alleviated and a risk of potential error exists on all levels. The risk of systematic error predominantly adheres to the fact that we could not blind the outcome assessors. Random error refers to the risk of type 1 and type 2 errors. The trial met the required sample size estimation both in regard to individuals and clusters. The large individual sample size dramatically reduces the risk of random error, however the number of clusters is equally important in a cluster randomised trial (CRT) and we just met the required number of clusters. The risk of design error primarily corresponds to the fact that it was not ethically feasible to dictate compliance with the predictions of the SARI model, i.e. the anaesthesiologist could chose to disregard the prediction comprised in the risk model. Moreover, the clustered design poses challenges in regards to unit of analysis and statistical adjustments. Finally, the primary outcome 'unanticipated difficult intubation' is merely a surrogate for morbidity and mortality, and even though 48 hours- and 30 days mortality were assessed they were secondary outcomes. However, we have strived to address the 'hardest' outcomes possible and believe that the primary outcome is in concordance with the GRADE category of outcomes, 'critical for decision making' [68].

CONCLUSION

The proportion of unanticipated difficult airway management is high in Denmark. From 2008 to 2011, 75 to 93% of all difficult intubations were unanticipated and a similar pattern was found for difficult mask ventilation.

We were not able to induce a reduction in our primary outcome, the incidence of unanticipated difficult intubations, by undertaking a large randomised multicentre trial and implementing preoperative use of the SARI compared to usual care. Although the unadjusted sensitivity and positive predictive value did increase in SARI departments compared to usual care departments in population 2, no statistical significant difference was found when adjusting for cluster and stratum affiliation. In comparison to Paper 1 (75-93%) the crude percentage of unanticipated difficulties in all difficult intubations, reduced in the DIFFICAIR trial to 45-89% in SARI departments and 60-91% in usual care departments. Although indications of improvement, these are predominantly found in population 2 (named sensitivity analysis 1 in Paper 1) and the extensive implementation of advanced equipment (e.g. video laryngoscopes) probably accounts for the majority of this effect. Nevertheless, the proportions of unanticipated airway difficulties found in this thesis, underline the continued challenge anaesthesiologists' face in predicting these events.

Clinical implications and perspectives

No other adequately powered randomised clinical trial has prospectively compared two different strategies for preoperative airway assessment and this thesis contributes to enhancing the understanding of airway related risks and difficulties. Over the time of this thesis, the attention to preoperative airway assessment has been heightened in Denmark. Our data may indicate a small increase in the predictive accuracy and a tendency towards enhanced allocation of resources to potential risk patients from Paper 1 to Paper 4. The level of airway assessment, for example the number of preoperative tests, appears to be quite good in Denmark, although no formal comparison has been made across borders. The intervention in Paper 4, did not prove to be efficient compared with the existing level of airway assessment in Denmark. However, this does not mean that every kind of airway assessment is equally good (or bad), nor that the intervention could not have potential benefits if compared to a 'usual care level' different than the Danish. Nevertheless, we have no well-founded reason to recommend the SARI model as a compulsory and superior approach to preoperative airway assessment compared to usual care based on the DIFFICAIR trial.

The SARI has now become recommended for preoperative airway assessment in several departments; introduced in the chapter on preoperative assessment in a textbook on basis anaesthesia; and incorporated in the formal education of anaesthesia specialists in the capital region [69]. This is based on the assumption that the SARI is a superior tool for airway assessment, something we were not able to demonstrate. However, one can hope that the tendency towards national systemisation and uniformity may have a positive impact for future patients.

METHODOLOGICAL PERSPECTIVES

Criteria for recommending the implementation of a predictive model in clinical practice

The desire to predict a future outcome from one or several patient related prognostic factors is fundamental in medicine. Good outcome prediction can alter and stratify the treatment for the individual patient and potentially improve the outcome. Good prediction of an outcome is rarely derived from a single factor and multiple factors build into a predictive model is often required in order to get adequate diagnostic accuracy. Optimally, the model produces an absolute risk of a certain outcome, however most commonly a model will estimate a relative risk. Good development and implementation of a predictive model is undertaken in four steps: 1) Estimation and quantification of a baseline risk or potential problem, e.g. finding a high proportion of unanticipated difficult airway management (Paper 1), 2) identification of potential risk factors and model development, e.g. building the SARI model [27], 3) Validation of the model in a external cohort, e.g. re-testing the diagnostic accuracy of the SARI in an independent cohort other than the one it was developed from, and ultimately 4) testing the clinical impact of the model in a comparative study versus usual care practice, e.g. testing the SARI in a randomised setting versus usual care (Paper 4).

Many predictive tools have been proposed for preoperative identification of patients at risk of a difficult intubation [27–30, 70]. Some of these tools may have been implemented in clinical practice and are therefore accepted as good predictive tools [43]. Unfortunately, none have been sufficiently validated or prospectively tested in a relevant clinical setting. Premature implementation of predictive or diagnostic tools is common and by no means an isolated anaesthesiological issue. It is not rare that a predictive tool finds its way into clinical practice based on step 2) development of a new model, showing promising good prediction. Inter-

nal bootstrap validation has become increasingly employed, but rarely is a predictive model tested in a independent cohort and comparative clinical impact studies is almost non-existent [33, 39, 71]. There are several potential pitfalls related to implementing predictive models into clinical practice without prior external validation or test of the clinical impact. Most importantly is the risk of overestimating the diagnostic accuracy of the model [33, 72]. Furthermore, there is a risk of extrapolating the model to a wider or deferent patient population than the one the model was developed in, without knowing the potential of the model in the new population [39, 72, 73].

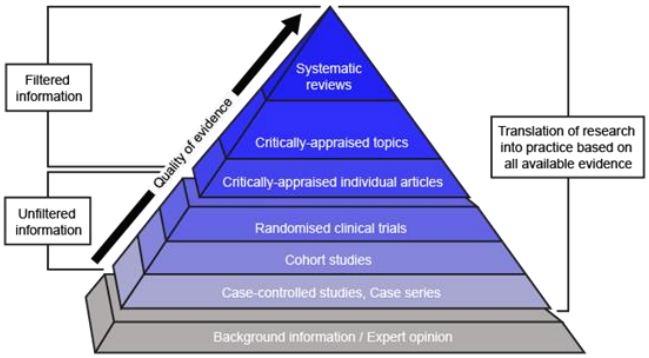
In 2013 the UK based PROGRESS group proposed a guideline for developing, validating and testing the clinical impact of prognostic models [33, 37, 60, 74]. One of the conclusions from the PROGRESS groups was that "researchers should shift to validation, updating, and impact studies of existing models". The SARI model has never been externally validated, however the individual risk factors and various combinations of the risk factors comprising the SARI has been validated in different cohorts [11]. The PROGRESS group further concluded that "clinical practice guideline recommendations relating to the use of prognostic models should be based on such impact studies" [33]. With the papers constituting this thesis we wanted to bring research on prediction of difficult airway management one step further by assessing the clinical impact of a predictive model.

The cluster randomised trial

The methodological advances of applying at well conducted randomised trial setting on reducing the risk of systematic error and confounding has been acknowledged for decades. The randomised clinical trial (RCT) therefore stands as the gold standard when comparing healthcare interventions [61]. However, it can be a challenging and costly task to conduct a well-powered RCT. Especially when the trial is addressing important severe adverse outcomes with a low event rate, requiring large numbers of patients. Observational studies can be conducted on large cohorts of patients, thus allowing detection of associations between an exposure and a rare, but severe adverse outcome. However, inherent in the observational design is the risk of several types of confounding [75]. Nevertheless, the limited feasibility of some RCTs have resulted in clinical recommendations based on lower levels of evidence, e.g. observational cohort studies.

When assessing the clinical impact of a predictive model on a relevant patient outcome, a comparative study is required. Two groups need to be compared: one using usual care and one using the model to guide treatment decisions. The scientifically strongest design for this comparison is the (cluster) randomised trial [33]. That being said, the cluster randomised trial has some methodological challenges. It is more prone to baseline imbalance, and thus residual confounding, compared to the individually low biased randomised clinical trial. The reliability of conclusions from a CRT probably range somewhere between the reliability of conclusions from a cohort study and the individually randomised clinical trial (Figure 9). There are several key arguments for randomisation by clusters [76, 77]: 1) the intervention is intended and delivered to all – or a large portion of – the people in a particular cluster of people (e.g. a new strategy for airway assessment), 2) the intervention is targeted at health professionals in order to improve a certain patient related outcome (e.g. education in the use of the SARI),

Figure 9. Quality of evidence (Source: www.ctu.dk)



3) the intervention is assessed at individual level, but the risk of contamination from the intervention to the control group is inevitable within the cluster. For example, it is impossible to dictate the anaesthesiologists to forget the SARI model when facing a patient randomised to receive usual airway assessment. Testing the implementation of a new guideline is therefore preferably done at a departmental (cluster) level in a CRT [33, 77]. Additionally, the CRT may have the advantage of potentially including a larger number of patients, thus making the trial logistically feasible and providing sufficient power in order to address low frequency outcomes in a randomised a 'low biased' setting. Having decided on randomisation at a higher level than the individual patient, e.g. at physician or departmental level, several considerations must be addressed in the design of the trial, and the analysis of data. The individuals within each cluster will inevitably be more correlated on outcome than individuals from different clusters. This may be due to patient demographics; differences in treatment standards; differences in adjuvant interventions; individual provider preferences etc. When performing the sample size estimation, it is therefore imperative to consider the within and between cluster variation [78–81]. It can be very difficult to quantify such a priori variations on the primary outcome between patients within the same cluster and between clusters. Optimally, baseline data from a period close to trial initiation are available on the primary outcome from the relevant clusters. In this thesis baseline DAD data (from Paper 1) allowed for appropriate sample size estimation prior to randomisation for the CRT (Paper 2). When performing a sample size estimation, it is a valid rule of thumb that increasing the number of clusters is far more potent in increasing power than an increment in the number of individuals within clusters, since the latter approaches a ceiling effect rather fast [79]. Due to intra cluster correlation and since the unit of randomisation is the cluster, whereas the unit of measurement is the individual patient, risk of imbalances is greater in the CRT than in the traditional RCT – especially when number of clusters are limited [77]. It is generally recommended to use some form of stratification in order to alleviate this potential imbalance and enhance power [37, 77, 79]. Clusters can be divided into different strata based on predefined baseline characteristics associated with the outcome (confounders); cluster size (when this is uneven); or/and (optimally) the primary outcome at baseline. Adherence to strata is then evenly balanced between the intervention groups striving for a good and even randomisation (Figure 8).

It is generally accepted that analyses of RCTs must be adjusted for potentially confounding covariates [82]. This is also applicable for

CRTs. However, the analyses of CRTs comprise further complexity, since it is difficult to estimate and adjust for the effect of the clustering. Nonetheless, it is important to employ statistical modelling that enables an adjustment for the cluster variable, and several models has been proposed, depending of the nature of the clustering [49, 83]. Likewise, adjustment for a stratification variable can be preferable. In the DIFFICAIR trial the odds ratios on the primary outcomes were adjusted for the cluster variable and stratum in a generalised estimating equation [49, 54, 84, 85].

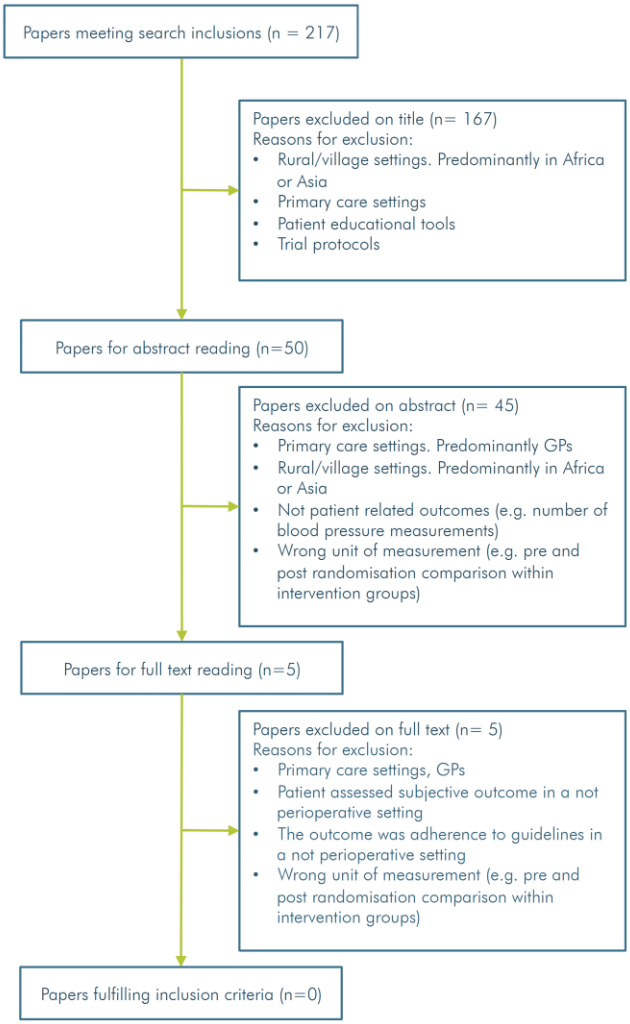
Ethical concerns have been raised with regard to informed consent in cluster randomised trials. Since there are two levels of inclusion (the cluster level and the individual patient level), yet one level of randomisation (the cluster level), the administrating authority (e.g. the department Head) accepts trial participation on behalf of all individuals in the cluster (e.g. the patients) [76, 86]. This may in some trial settings interfere with the ethics of individual patient consent for participating in a clinical trial. In the DIFFICAIR trial we did not dictate a certain approach for airway management of the patients and The Committee on Health Research Ethics of the Capital Region of Denmark therefore regarded the implementation of the intervention as a quality insurance project. Thus, individual patient consent was exempted.

Methodology of the PhD in the context of medical research

We sought to employ state of the art methodology for testing the implementation of a recommendation, when conducting the work comprised in this thesis. Initially, we conducted a baseline study on the proportion of unanticipated difficult intubation in Denmark. Using these data, we were able to identify a clinical problem and a potential for improvement. Further, baseline data allowed for appropriate sample size estimation for a cluster randomised trial. Sample size estimations were adjusted for between cluster variance on the primary outcome and this data additionally allowed for stratification on the primary outcome. Data analyses were conducted using appropriate statistical modelling and adjustment.

More and more prognostic models are being developed, yet they are scarcely tested on their impact in clinical practice. In a systematic review made by the PROGRESS group they identified only two published analyses from 2006 to 2009 on the impact of a prognostic model and when including previous reviews only ten such publications were identified [33]. These papers were not necessarily on perioperative or in-hospital predictive or prognostic models. We found it interesting to elucidate, whether state of the art methodology for testing the clinical impact of a prognostic model or recommendation had been applied before in a perioperative setting. A systematic MEDLINE search was conducted including all publications addressing new recommendations or guidelines in a cluster randomised setting. Inclusion criteria were cluster randomised trials testing a recommendation or prognostic model in a perioperative setting on a patient related outcome. The search strategy included all spellings and combinations of "cluster randomised trial" and was combined with terms regarding recommendations, guidelines or usual care/standards. Papers with titles referring to trial protocols were excluded in the search. The search resulted in 217 hits. The number was brought down to 50 papers after a read-through of the titles. Five papers were left for full text read after reading the 50 abstracts. None of the papers fulfilled the inclusion criteria after reading the final five papers (Figure 10). The vast majority of the excluded trials were

Figure 10. Inclusion strategy for the review on papers employing cluster randomised methodology and testing implementation of a recommendation



conducted in rural settings, e.g. having villages in Africa as the level of clustering, or conducted in primary care with the general practitioner as the most common level of clustering. Some trials investigated patients' educational tools, e.g. cell phone applications for diabetic control, and several did not measure patient related outcome, but merely tested the level of registration of the recommendation. Albeit, the search strategy may not have been completely exhaustive, the DIFFICAIR trial appear to be the first cluster randomised trial testing the implementation of a guideline in a perioperative setting.

The 50 abstracts were further investigated in order to quantify if any of the trials, regardless of cluster settings, had been able to demonstrate an effect of the intervention on a patient related outcome. Several trials had been able to show that the intervention led to better adherence to guidelines; changes in 'risk profile'; reduction in prescription of antibiotics; or enhanced use of testing.

It was encouraging to observe that several trials demonstrated better adherence to guidelines when the providers where taught and encouraged in the use of the intervention. This may support the assumed value of developing educational tools (e.g. a video and a white coat aid) and doing repeated teaching of the SARI in

the DIFFICAIR trial. However, very few trials were able to demonstrate effects of the intervention on patient related outcomes such as mortality, adverse events or even surrogates as e.g. blood pressure levels. Only one paper was able to present an intervention effect on a patient relevant outcome, reducing hospital admissions and mortality through the use of telehealth devices versus usual care [87].

Implications for future research

The DIFFICAIR trial provided information on more than one hundred thousand patients, and this information needs to be explored further. More than 22,000 patients were intubated following a complete SARI registration, and to aim for external validation and updating of the SARI model seems reasonable. Denmark is world renowned for its many comprehensive and high quality registries and databases. Valuable patient information (e.g. from perioperative or intensive care settings) and important patient related outcomes are recorded. As predictive models are becoming abundant in medical literature, still very few are tested for real clinical impact. To use a national clinical database as the platform for testing the implementation of a new recommendation in a randomised trial setting is innovative and may prove useful to others. This thesis poses an example of, how to test the implementation of a predictive model using a cluster randomised design. It is our hope that the methodology can serve as a precedent for testing and facilitating implementation of evidence-based recommendations.

Likewise, variables or potential risk factors registered in clinical databases need to be based on evidence, and improved methodology for CRTs may contribute to evidencebased development and evolvement of clinical databases.

ABBREVIATIONS

ASA	American Society of Anesthesiologists
BMI	Body Mass Index
CI	Confidence Interval
CRT	Cluster Randomised Trial
DAD	Danish Anaesthesia Database
GEE	Generalised Estimating Equation
ITT	Intention To Treat
NA	Not Applicable
NAP4	Fourth National Audit Project
NNT	Number Needed to Treat
OR	Odds Ratio
RCT	Randomised Clinical Trial
SARI	Simplified Airway Risk Index
UK	United Kingdom

DEFINITIONS

Advanced intubation techniques	In DAD, and in this thesis, defined as techniques for tracheal intubation that are more advanced than a conventional laryngoscope, e.g. video laryngoscope or fibre optic scope.
Between cluster variance	The variance in means, rates or proportions of an outcome between clusters.
Design error	Error resulting from applying the wrong design to answer a given clinical question (or vice versa).
Generalised estimating equation	A statistical model used to adjust for clustering in the data (certain observations being more correlated than others). The model accounts for intra cluster correlations on the outcome using a correlation matrix. In the DIFFICAIR trial (as recommended for this type of trial) we used an exchangeable correlation matrix, assuming equal correlation between any pair of observations within a cluster.
False positives	Patients who tested (incorrectly) positive, but did not experience the event.
Intention to treat analysis	Analysis based on the initial 'treatment' assignment (e.g. receiving SARI assessment), not the actual 'treatment' received (e.g. receiving complete or incomplete SARI assessment). There is no good consensus on handling of missing data in intention to treat analysis. In the DIFFICAIR trial we applied multiple imputation on missing values of the SARI score.
Intra cluster correlation	The ratio of between cluster variance to the total variance (Between cluster variation/ (Between cluster variation + Within cluster variation)). It ranges between 0 and 1; 0 meaning no variation between clusters.
Multiple Imputation	Statistical method for estimating missing values for any variable. The missing values are replaced with imputed values that are generated based on existing values from other variables. This results in a full data set (imputed dataset). Multiple imputed datasets are generated and are then combined to produce a pooled analysis result.
Negative likelihood ratio	Estimates how much the odds of experiencing the event decrease when the test is negative.
Negative predictive value	The proportion of patients who tested negative and who were correctly diagnosed as such.
P value	The probability of obtaining a result equal to, or even more extreme, than the one observed, under the assumption of the null hypothesis being true.
Positive likelihood ratio	Estimates how much the odds of experiencing the event increase

	when the test is positive.
Positive predictive value	The proportion of patients who tested positive and who were correctly diagnosed as such.
Random error	Error resulting from 'play of chance' i.e. drawing a false conclusion based on sparse data. Two types of false conclusions (errors) exist: type 1 and type 2 errors.
Sensitivity	The proportion of positives (patients experiencing an event), correctly identified by the test.
Specificity	The proportion of negatives (patients not experiencing an event), correctly identified by the test.
Systematic error	Error resulting from methodological conduct causing an increase in the risk of drawing an erroneous conclusion. Also called bias.
True positives	Patients who tested (correctly) positive, and subsequently experienced the event.
Type 1 error	Incorrect rejection of the null hypothesis.
Type 2 error	Incorrect rejection of an alternative hypothesis.
Within cluster variation	The variance of an outcome between individuals within a cluster.

SUMMARY

Difficulties with airway management in relation to general anaesthesia have been a challenge for the anaesthesiologist since the birth of anaesthesia. Massive landmark improvements have been made and general anaesthesia is now regarded as a safe procedure. However rare, difficult airway management still occurs and it prompts increased risk of morbidity and mortality – especially when not anticipated. Several preoperative risk factors for airway difficulties have been identified, yet none have convincing diagnostic accuracy as stand alone tests. Combining several risk factors increase the predictive value of the test and multivariable risk models have been developed. The 'Simplified Airway Risk Index' (SARI) is a predictive model developed for anticipation of a difficult direct laryngoscopy. However, neither the diagnostic accuracy of the SARI nor of any other model has been tested prospectively and compared with existing practice for airway assessment in a randomised trial setting.

The first objective of this thesis was to quantify the proportion of unanticipated difficult intubation and difficult mask ventilation in Denmark.

The second objective was to design a cluster randomised trial, using state of the art methodology, in order to test the clinical impact of using the SARI for preoperative airway assessment compared with a clinical judgement based on usual practice for airway assessment.

Finally, to test if implementation of the SARI would reduce the proportion of unanticipated difficult intubation compared with usual care for airway assessment.

This thesis is based on data from the Danish Anaesthesia Database (DAD). Paper 1 presents an observational cohort study on 188,064 patients who underwent tracheal intubation from 2008 to 2011. Data on the anaesthesiologists' preoperative anticipations of airway difficulties was compared with actual airway management conditions, thus enabling an estimation of the proportion of unanticipated difficulties with intubation and mask ventilation.

Papers 2 and 3 outline the methodology and the pre-trial calculations and considerations leading to the DIFFICAIR trial described in Paper 4. The trial was designed to randomise anaesthesia department to either thorough education in, and subsequent use of the SARI for preoperative airway assessment or to continue usual care. Registration of the SARI in DAD was made mandatory in SARI departments and impossible in usual care departments. Conditions regarding anticipation of difficulties and actual airway managements were recorded as for Paper 1. DAD data made it possible to estimate an appropriate sample size, considering the between cluster variation, and to construct a stratification variable based on 2011 baseline values of the primary outcome used in the DIFFICAIR trial.

Paper 1 revealed that 1.86% of all patients who were intubated, but not planned for advanced intubation techniques (e.g. video laryngoscopy), were unanticipated difficult to intubate. However, 75 to 93% of all difficult intubations were unanticipated. Furthermore, 94% of all difficult mask ventilations were unanticipated. In Paper 4, 59,514 patients were included in the primary analyses. The proportion of unanticipated difficult intubations was 2.38% (696/29,209) in SARI departments and 2.39% (723/30,305) in usual care departments. The adjusted odds ratio was 1.03 (95% CI: 0.77–1.38), P = 0.84. No significant differences were detected in other adjusted outcome measures and neither a 58% increase in patients anticipated to have intubation difficulties nor an 84% increase in patients scheduled for advanced intubation techniques in SARI departments reached statistical significance, P = 0.29 and P = 0.06 respectively.

The papers constituting this thesis demonstrate that at high proportion of airway management difficulties are unanticipated. In a cluster randomised trial it was not possible to reduce the proportion of unanticipated difficult intubation in daily clinical practice by implementing a systematic approach for airway assessment compared with usual care. However, implementation of the SARI may increase the anticipation of intubation difficulties and it may change practice towards advanced intubation techniques. This thesis underlines the continued challenge anaesthesiologists face in predicting airway management related difficulties.

REFERENCES

1. Cook T, Woodall N, Frerk C: Major complications of airway management in the United Kingdom. Natl Audit Proj 2011.
2. Woodall NM, Cook TM: National census of airway management techniques used for anaesthesia in the UK: first phase of the Fourth National Audit Project at the

Royal College of Anaesthetists. Br J Anaesth 2011, 106:266–71.
3. Hove LD, Steinmetz J, Christoffersen JK, Møller A, Nielsen J, Schmidt H: Analysis of deaths related to anesthesia in the period 1996-2004 from closed claims registered by the Danish Patient Insurance Association. Anesthesiology 2007, 106:675–80.
4. McClure JH, Cooper GM, Clutton-Brock TH: Saving mothers' lives: reviewing maternal deaths to make motherhood safer: 2006-8: a review. Br J Anaesth 2011, 107:127–32.
5. Peterson GN, Domino KB, Caplan RA, Posner KL, Lee LA, Cheney FW: Management of the difficult airway - A closed claims analysis. Anesthesiology 2005, 103:33–39.
6. Rosenstock C, Møller J, Hauberg a: Complaints related to respiratory events in anaesthesia and intensive care medicine from 1994 to 1998 in Denmark. Acta Anaesthesiol Scand 2001, 45:53–8.
7. Kheterpal S, Martin L, Shanks AM, Tremper KK: Prediction and outcomes of impossible mask ventilation: a review of 50,000 anesthetics. Anesthesiology 2009, 110:891–897.
8. Ramachandran SK, Mathis MR, Tremper KK, Shanks AM, Kheterpal S: Predictors and clinical outcomes from failed Laryngeal Mask Airway UniqueTM: a study of 15,795 patients. Anesthesiology 2012, 116:1217–26.
9. Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Gätke MR, Wetterslev J: A documented previous difficult tracheal intubation as a prognostic test for a subsequent difficult tracheal intubation in adults. Anaesthesia 2009, 64:1081–8.
10. Kheterpal S, Healy D, Aziz MFM, Shanks AM, Freundlich RE, Linton F, Martin LD, Linton J, Epps JL, Fernandez-Bustamante A, Jameson LC, Tremper T, Tremper KK: Incidence, Predictors, and Outcome of Difficult Mask Ventilation Combined with Difficult Laryngoscopy: A Report from the Multicenter Perioperative Outcomes Group. Anesthesiology 2013, 119:1360–9.
11. Shiga T, Wajima Z, Inoue T, Sakamoto S, Sakamoto A: Predicting difficult intubation in apparently normal patients. Anesthesiology 2005, 103:429–437.
12. Kheterpal S, Han R, Tremper KK, Shanks AM, Tait AR, O'Reilly M, Ludwig TA, Martin L: Incidence and predictors of difficult and impossible mask ventilation. Anesthesiology 2006, 105:885–891.
13. Nørskov AK, Rosenstock C V, Wetterslev J, Astrup G, Afshari a, Lundstrøm LH: Diagnostic accuracy of anaesthesiologists' prediction of difficult airway management in daily clinical practice: a cohort study of 188 064 patients registered in the Danish Anaesthesia Database. Anaesthesia 2014:1–10.
14. Adnet F, Borron SW, Racine SX, Clemessy J-L, Fournier J-L, Plaisance P, Lapandry C: The intubation difficulty scale (IDS). Proposal and evaluation of a new score charachterizing the complexity of endotracheal intubation. Anesthesiology 1997, 87:1290–1297.
15. Law JA, Broemling N, Cooper RM, Drolet P, Duggan L V, Griesdale DE, Hung OR, Jones PM, Kovacs G, Massey S, Morris IR, Mullen T, Murphy MF, Preston R, Naik VN, Scott J, Stacey S, Turkstra TP, Wong DT: The difficult airway with recommendations for management--part 1--difficult tracheal intubation encountered in an unconscious/induced patient. Can J Anaesth 2013, 60:1089–118.
16. Huitink JM, Bouwman R: The myth of the difficult airway: airway management revisited. Anaesthesia 2014, e pub.
17. Apfelbaum JL, Hagberg CA, Caplan RA, Blitt CD, Connis RT, Nickinovich DG, Benumof JL, Berry FA, Bode RH, Cheney FW, Guidry OF, Ovassapian A: Practice guidelines for management of the difficult airway: an updated report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. Anesthesiology 2013, 118:251–70.
18. Kheterpal S, Han R, Tremper KK, Shanks A, Tait AR, O'Reilly M, Ludwig T a: Incidence and predictors of difficult and impossible mask ventilation. Anesthesiology 2006, 105:885–91.
19. Langeron O, Masso E, Huraux C, Guggiari M, Bianchi A, Coriat P, Riou B: Prediction of difficult mask ventilation. Anesthesiology 2000, 92:1229–1236.
20. Cormack RS, Lehane J: Difficult tracheal intubation in obstetrics. Anaesthesia 1984, 39:1105–11.
21. Han R, Tremper K, Kheterpal S, O'Reilly M: Grading scale for mask ventilation. Anesthesiology 2004, 101:267.
22. Cook TM, Macdougall-Davis SR: Complications and failure of airway management. Br J Anaesth 2012, 109(Supplement 1):i68–i85.
23. www.ncepod.org.uk
24. Cook TM, Woodall N, Frerk C: Major complications of airway management in the UK: results of the Fourth National Audit Project of the Royal College of Anaesthetists and the Difficult Airway Society. Part 1: anaesthesia. Br J Anaesth 2011, 106:617–31.
25. Caplan R, Benumof J, Berry F: Practice guidelines for management of the difficult airway: an updated report by the American Society of Anesthesiologists Task Force on Management of the Difficult Airway. Anesthesiology 2003, 98:1269–1277.
26. Yentis SM: Predicting difficult intubation--worthwhile exercise or pointless ritual?. Anaesthesia 2002, 57:105–9.
27. El-Ganzouri AR, McCarthy RJ, Tuman KJ, Tanck EN, Ivankovich AD: Preoperative airway assessment: predictive value of a multivariate risk index. Anesth Analg 1996, 82:1197–1204.
28. L'Hermite J, Nouvellon E, Cuvillon P, Fabbro-Peray P, Langeron O, Ripart J: The Simplified Predictive Intubation Difficulty Score: a new weighted score for difficult airway assessment. Eur J Anaesthesiol 2009, 26:1003–9.
29. Naguib M, Scamman FL, O'Sullivan C, Aker J, Ross AF, Kosmach S, Ensor JE: Predictive performance of three multivariate difficult tracheal intubation models: a double-blind, case-controlled study. Anesth Analg 2006, 102:818–24.
30. Arné J, Descoins P, Fusciardi J, Ingrand P, Ferrier B, Boudigues D, Ariès J: Preoperative assessment for difficult intubation in general and ENT surgery: predictive value of a clinical multivariate risk index. Br J Anaesth 1998, 80:140–6.
31. O'Sullivan E, Laffey J, Pandit JJ: A rude awakening after our fourth "NAP": lessons for airway management. Anaesthesia 2011, 66:331–334.
32. Deeks JJ, Altman DG: Diagnostic tests 4: likelihood ratios. BMJ 2004, 329:168–9.

33. Steyerberg EW, Moons KGM, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG: Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 2013, 10:e1001381.
34. Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Wetterslev J: High body mass index is a weak predictor for difficult and failed tracheal intubation: a cohort study of 91,332 consecutive patients scheduled for direct laryngoscopy registered in the Danish Anesthesia Database. Anesthesiology 2009, 107:266–274.
35. Lundstrøm LH, Vester-Andersen M, Møller AM, Charuluxananan S, L'hermite J, Wetterslev J: Poor prognostic value of the modified Mallampati score: a meta-analysis involving 177 088 patients. Br J Anaesth 2011, 107:659–67.
36. Lee A, Fan LTY, Gin T, Karmakar MK, Ngan Kee WD: A systematic review (meta-analysis) of the accuracy of the Mallampati tests to predict the difficult airway. Anesth Analg 2006, 102:1867–1878.
37. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Kyzas PA, Malats N, Briggs A, Schroter S, Altman DG, Hemingway H, Kyzas PA: Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS Med 2013, 10:e1001380.
38. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJCR, Vergouwe Y, Habbema JDF: Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001, 54:774–781.
39. Moons KGM, Altman DG, Vergouwe Y, Royston P: Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009, 338:b606.
40. Samsoon GLT, Young JRB: Difficult tracheal intubation: a retrospective study. Anaesthesia 1987, 42:487–490.
41. Mallampati SR, Gatt SP, Gugino LD, Desai SP, Waraksa B, Freiberger D, Liu PL: A clinical sign to predict difficult tracheal intubation: a prospective study. Can Anaesth Soc J 1985, 32:429–34.
42. Altman DGD, Royston P: The cost of dichotomising continuous variables. BMJ 2006, 332:1080.
43. Nørskov AK, Rosenstock CV, Wetterslev J, Lundstrøm LH: Preoperative airway assessment. No national consensus in Denmark. Eur J Anaesthesiol 2014, 31(e-suppl 52, Abstract programme):275.
44. Vester-Andersen M, Lundstrøm LH, Møller MH, Waldau T, Rosenberg J, Møller a M: Mortality and postoperative care pathways after emergency gastrointestinal surgery in 2904 patients: a population-based cohort study. Br J Anaesth 2014, 112:860–70.
45. Lund C a, Møller AM, Wetterslev J, Lundstrøm LH: Organizational factors and long-term mortality after hip fracture surgery. A cohort study of 6143 consecutive patients undergoing hip fracture surgery. PLoS One 2014, 9:e99308.
46. Chan A, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin JA, Doré CJ, Parulekar WR, Summerskill WSM, Groves T, Schulz KF, Sox HC, Rockhold FW, Rennie D, Moher D: SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013, 158:200–7.
47. Dwan K, Gamble C, Williamson PR, Altman DG: Reporting of clinical trials: a review of research funders' guidelines. Trials 2008, 9:66.
48. Fergusson D, Aaron SD, Guyatt G, Hébert P: Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ 2002, 325:652–4.
49. Klar N, Donner A: Current and future challenges in the design and analysis of cluster randomization trials. Stat Med 2001, 20:3729–3740.
50. Hayes RJ, Moulton LH: Cluster Randomised Trials (Chapman & Hall/CRC Interdisciplinary Statistics). Chapman and Hall/CRC; 2009:338.
51. Kahan BC, Morris TP: Assessing potential sources of clustering in individually randomised trials. BMC Med Res Methodol 2013, 13:58.
52. Grimshaw JM, Eccles MP: Is evidence-based implementation of evidence-based care possible?. Med J Aust 2004, 180(6 Suppl):S50–1.
53. Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L, Wensing M, Dijkstra R, Donaldson C: Effectiveness and efficiency of guideline dissemination and implementation strategies. Health Technol Assess (Rockv) 2004, 8:iii–iv, 1–72.
54. Kahan BC, Morris TP: Improper analysis of trials randomised using stratified blocks or minimisation. Stat Med 2012, 31:328–40.
55. Peters TJ, Richards SH, Bankhead CR, Ades AE, Sterne JAC: Comparison of methods for analysing cluster randomized trials: an example involving a factorial design. Int J Epidemiol 2003, 32:840–846.
56. Ma J, Raina P, Beyene J, Thabane L: Comparison of population-averaged and cluster-specific models for the analysis of cluster randomized trials with missing binary outcomes: a simulation study. BMC Med Res Methodol 2013, 13:9.
57. Randell T: Prediction of difficult intubation. Acta Anaesthesiol Scand 1996, 40:1016–1023.
58. Colli A, Fraquelli M, Casazza G, Conte D, Nikolova D, Duca P, Thorlund K, Gluud C: The architecture of diagnostic research: from bench to bedside--research guidelines using liver stiffness as an example. Hepatology 2014, 60:408–18.
59. Royston P, Altman DG, Sauerbrei W: Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006, 25:127–41.
60. Hingorani ADA, Windt DA van der A van der, Riley RD, Abrams K, Moons KGM, Steyerberg EW, Schroter S, Sauerbrei W, Altman DG, Hemingway H: Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ 2013, 5793:1–9.
61. Schulz KF, Altman DG, Moher D: CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 2010, 7:e1000251.
62. Nørskov AK, Rosenstock CV, Wetterslev J, Lundstrøm LH: Incidence of unanticipated difficult airway using an objective airway score versus a standard clinical airway assessment: the DIFFICAIR trial – trial protocol for a cluster randomized clinical trial. Trials 2013, 14:347.
63. Nørskov AK, Lundstrøm LH, Rosenstock CV, Wetterslev J: Detailed statistical analysis plan for the difficult airway management (DIFFICAIR) trial. Trials 2014, 15:173.

64. Higgins JPT, Green S: Cochrane Handbook for Systematic Reviews of Interventions. Volume 5.0.1; 2008:672.
65. Jakobsen JC, Gluud C, Winkel P, Lange T, Wetterslev J: The thresholds for statistical and clinical significance - a five-step procedure for evaluation of intervention effects in randomised clinical trials. BMC Med Res Methodol 2014, 14:34.
66. Keus F, Wetterslev J, Gluud C, van Laarhoven CJHM: Evidence at a glance: error matrix approach for overviewing available evidence. BMC Med Res Methodol 2010, 10:90.
67. Nørskov AK, Lundstrøm LH, Rosenstock CV, Wetterslev J: Anticipation of the difficult airway. Br J Anaesth 2014, 112:2014.
68. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schünemann HJ: What is "quality of evidence" and why is it important to clinicians?. BMJ 2008, 336:995–8.
69. Rasmussen LS, Steinmetz J: Anæstesi, 4. Udgave, FADLs Forlag. 2014.
70. Wilson ME, Spiegelhalter D, Robertson JA, Lesser P: Predicting Difficult Intubation. Br J Anaesth 1988, 61:211–216.
71. Gluud C, Gluud LL: Evidence based diagnostics. BMJ 2005, 330:724–6.
72. Moons KGMK, Royston P, Vergouwe Y, Grobbee DE, Altman DG: Prognosis and prognostic research: what, why, and how?. BMJ 2009, 338:b375.
73. Altman DG, Vergouwe Y, Royston P, Moons KGM: Prognosis and prognostic research: validating a prognostic model. BMJ 2009, 338:b605.
74. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, Briggs A, Udumyan R, Moons KGM, Steyerberg EW, Roberts I, Schroter S, Altman DG, Riley RD, for the Progress Group: Prognosis research strategy (PROGRESS) 1: A framework for researching clinical outcomes. BMJ 2013, 346:e5595.
75. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG: Evaluating non-randomised intervention studies. Health Technol Assess 2003, 7:iii–x, 1–173.
76. Edwards SJL, Braunholtz DA, Lilford RJ, Stevens AJ: Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ 1999, 318:1407–9.
77. Medical Research Council: Cluster Randomised Trials: Methodological and Ethical Considerations. 2002.
78. Hayes RJ, Bennett S: Simple sample size calculation for cluster-randomized trials. Int J Epidemiol 1999, 28:319–26.
79. Campbell MK, Piaggio G, Elbourne DR, Altman DG: Consort 2010 statement: extension to cluster randomised trials. BMJ 2012, 345:e5661.
80. Eldridge SM, Ashby D, Kerry S: Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. Int J Epidemiol 2006, 35:1292–300.
81. Ahn C, Hu F, Skinner CCS, Ahn D: Effect of Imbalance and Intracluster Correlation Coefficient in Cluster Randomization Trials with Binary Outcomes When the Available Number of Clusters is Fixed in Advance. Contemp Clin Trials 2009, 30:317–320.
82. Hernández A V, Steyerberg EW, Habbema JDF: Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. J Clin Epidemiol 2004, 57:454–60.
83. Piaggio G, Carroli G, Villar J, Pinol A, Bakketeig L, Lumbiganon P, Bergsjø P, Al-Mazrou Y, Ba'aqeel H, Belizán JM, Farnot U, Berendes H: Methodological considerations on the design and analysis of an equivalence stratified cluster randomization trial. Stat Med 2001, 20:401–16.
84. Kahan BC, Morris TP: Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects?. Stat Med 2013, 32:1136–49.
85. Kahan BC, Morris TP: Adjusting for multiple prognostic factors in the analysis of randomised trials. BMC Med Res Methodol 2013, 13:99.
86. Donner A, Klar N: Pitfalls of and controversies in cluster randomization trials. Am J Public Health 2004, 94:416–22.
87. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Cartwright M, Rixon L, Knapp M, Henderson C, Rogers A, Fitzpatrick R, Hendy J, Newman S: Effect of telehealth on use of secondary care and mortality: findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012, 344:e3874.