FACULTY OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

PhD thesis

Pre-operative airway assessment - Experience gained from a multicentre cluster randomised trial and the Danish Anaesthesia Database

Anders Kehlet Nørskov, MD

Department of Anaesthesiology and Intensive Care Medicine

Nordsjællands Hospital, University of Copenhagen, Denmark

and

Copenhagen Trial Unit, Centre for Clinical Intervention Research

Rigshospitalet, University of Copenhagen, Denmark

2015

PhD thesis

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

Anders Kehlet Nørskov, MD

Department of Anaesthesiology and Intensive Care Medicine

Nordsjællands Hospital

Dyrehavevej 29, 3400 Hillerød

Denmark

Mail: anderskehlet@hotmail.com

and

Copenhagen Trial Unit, Centre for Clinical Intervention Research

Copenhagen University Hospital, Rigshospitalet

Blegdamsvej 9, 2100 København Ø

Denmark

This thesis was submitted to the Graduate School of The Faculty of Health and Medical Sciences, University of Copenhagen

Academic advisors

Lars Hyldborg Lundstrøm, Associate Professor, MD, PhD, BSc (Faculty advisor)
Department of Anaesthesiology and Intensive Care Medicine
Nordsjællands Hospital, University of Copenhagen
Denmark

Charlotte Vallentin Rosenstock, Associate Professor, MD, PhD (Project advisor)
Department of Anaesthesiology and Intensive Care Medicine
Nordsjællands Hospital, University of Copenhagen
Denmark

Jørn Wetterslev, Chief Physician, MD, PhD (Project advisor)
Copenhagen Trial Unit, Centre for Clinical Intervention Research, Dpt. 7812
Rigshospitalet, University of Copenhagen
Denmark

Evaluating committee

Anders Perner, Professor, MD, PhD (Chairperson)
Department of Intensive Care Medicine, Dpt. 4131
Rigshospitalet, University of Copenhagen
Denmark

Annmarie Touborg Lassen, Professor, MD, PhD, DMSc
Department of Emergency Medicine
Odense University Hospital, University of Southern Denmark
Denmark

Niklas Nielsen, Associate Professor, MD, PhD
Department of Clinical Sciences
Lund University
Sweden

Original papers

This thesis is based on the following papers:

I. 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

Nørskov AK, Rosenstock CV, Wetterslev J, Astrup G, Afshari A, Lundstrøm LH.

Anaesthesia 2015, **70**;272-281.

II. 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

Nørskov AK, Rosenstock CV, Wetterslev J, Lundstrøm LH.

Trials 2013, **14**:347.

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

Nørskov AK, Lundstrøm LH, Rosenstock CV, Wetterslev J.

Trials 2014, **15**:173.

IV. 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

Nørskov AK, Wetterslev J, Rosenstock CV, Afshari A, Astrup G, Jakobsen JC, Lundstrøm LH and the DIFFICAIR trial author group.

Anesthesiology, Submitted for publication.

Content

Preface

Looking back on the last couple of years, it has been an amazing run and I feel truly privileged. I have grown both academically and personally and the list of people I owe my gratitude to is long. I was fortunate enough to be welcomed into the most inspiring, caring and meticulous group of people, and it has been an immense pleasure receiving guidance and supervision, exceeding any expectations, from my three advisors. First of all, thanks to Lars for taking the role of my 'older brother', leading me safely through – at times – rough waters and allowing me to dodge many bullets. You have always had my interest in mind, you have encouraged me and you have become a friend.

I thank Charlotte for taking me under her wing and for always being there professionally as well as personally.

Jørn took a chance with me when I came knocking on his door and with his persistent, uncompromising and meticulous approach to research I could not have had a better mentor. I also want to thank Jørn for his inspiring lead in the endeavour of doing things right.

This PhD study was done in cooperation with the Department of Anaesthesiology at Nordsjælland Hospital and Copenhagen Trial Unit at Rigshospitalet. I thank the staff at both places for taking an interest in, and facilitating the process.

Special thanks to Kristian Antonsen as Head of the Department of Anaesthesiology at Nordsjælland Hospital in 2011, for believing in me as well as the project. Also, I would like to thank the present Head of the Department of Anaesthesiology, Lars Steen Jakobsen, for always making me feel as a part of the department, even though my official employment shifted to Rigshospitalet.

I thank Christian Gluud, Head of Copenhagen Trial Unit, for his encouraging spirit and for letting me become a part of an inspiring environment.

The TRYG Foundation has funded almost all of my work and without these funds this PhD study would have had a cumbersome course. My shoulders dropped when I was relieved of economical worries and allowed to focus on my goals instead.

The combined brain trust and timely idiocy of my fellow PhD students Charlotte, Semera, Christian R, Theis, Christian SH and Rasmus have made my days equally fun and inspiring. Thank you guys. I thank Anette J Madsen and Helle Sinkbæk for all your graphical magic and Mette Hansen for practical assistance.

This PhD would never have seen the light of day without the acceptance to participate in the DIFFICAIR trial from the involved department Heads, and I am greatly indebted to the many local investigators for a pivotal effort. I thank you all for your dedication to the project.

I am grateful to Grethe Astrup and Arash Afshari for accepting to take part in the DIFFICAIR steering committee and for excellent and prompt commenting on my work. I owe Janus C Jakobsen sincere thanks for his tireless approach to research and his statistical assistance.

I thank Jakob L Thomsen for his creative inspiration and support, Jasper Overgaard for his creativity, and Günther Momsen for practical assistance in wrestling the Danish Anaesthesia Database.

Lastly, I wish to thank Frederikke for always putting a smile on my face and my family and friends for simply always being there.

Anders Nørskov

January 2015

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 SA-

	RI 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

Background

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 pre-operative 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.

Objectives

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 pre-operative 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.

Methods

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' pre-operative 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 pre-operative 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.

Results

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.

Conclusion

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.

Dansk resumé

Baggrund

Vanskelig luftvejshåndtering i forbindelse med generel anæstesi har udfordret anæstesilægen så længe generel anæstesi har eksisteret. Fra den spæde begyndelse og til nu er der sket massive og skelsættende forbedringer og generel anæstesi betragtes nu som en sikker procedure. Omend sjældent, så opstår vanskelig luftvejshåndtering dog stadig og det øger risikoen for morbiditet og mortalitet - især når det ikke er forventet. Selvom adskillige risikofaktorer for vanskelig luftvejshåndtering er blevet identificeret, har ingen dog vist overbevisende diagnostisk præcision som enkeltstående test. Ved at kombinere flere risikofaktorer øges den prædiktive værdi af testen og flere multivariable risikomodeller er derfor blevet udviklet. "The Simplified Airway Risk Index" (SARI) er en prædiktiv model, der er udviklet til at forudsige vanskelig direkte laryngoskopi. Imidlertid er hverken den diagnostiske præcision af SARI modellen eller af nogen anden model blevet eftertestet prospektivt i en sammenligning med gængs praksis for luftvejsvurdering i et randomiseret forsøg.

Formål

Første formål med denne afhandling var at opgøre andelen af uventet vanskelig intubation og uventet vanskelig maskeventilation.

Det andet formål var at designe et cluster randomiseret forsøg, der byggede på "state of the art" metodologi til at teste den kliniske effekt af at implementere SARI modellen som instruks for præoperativ luftvejsvurdering mod en klinisk vurdering baseret på vanlig praksis på området ("Usual care").

Det endelige formål var at teste, om andelen af uventet vanskelig intubation kunne nedbringes ved at implementere SARI modellen som instruks for luftvejsvurdering i sammenligning med vanlig praksis.

Metode

Denne afhandling er baseret på data fra Dansk Anæstesi Database (DAD). Artikel (Paper) 1 præsenterer et observationelt kohorte studie på 188.064 patienter, som blev trakealt intuberet fra 2008 til 2011. Anæstesilægens præoperative forventning om luftvejsproblemer er sammenlignet med de faktiske luftvejshåndteringsforhold. Andelen af uventet vanskelig intubation og uventet vanskelig maskeventilation kunne hermed opgøres.

I Artikel 2 og 3 beskrives metoden, samt de beregninger og overvejelser, der førte til DIFFICAIR forsøget, som er beskrevet i Artikel 4. Forsøget var designet til at randomisere anæstesiafdelinger til enten: grundig undervisning i- og brug af SARI modellen til præoperativ luftvejsvurdering eller til at fortsætte med vanlig procedure. Registrering af SARI blev gjort obligatorisk i DAD på SARI afdelingerne, men skjult og ikke mulig at registrere på "Usual care" afdelinger. Forventninger om luftvejsproblemer og faktiske luftvejshåndterings forhold blev registreret som for Artikel 1. DAD data gjorde det muligt at foretage en sample size estimation, der bl.a. tog højde for "between cluster" variation, samt at konstruere en stratifikationsvariabel baseret på 2011 baseline værdier af det primære effektmål, der anvendes i DIFFICAIR forsøget.

Resultater

Artikel 1 viste, at 1,86% af alle intuberede patienter, der ikke var planlagt til avanceret intubationsteknik (f.eks. videolaryngoskopi), var uventet vanskelige at intubere. 75 til 93% af alle vanskelige intubationer var imidlertid ikke forudset. Endvidere var 94% af alle vanskelige maskeventilationstilfælde uventet vanskelige. I Artikel 4 blev 59.514 patienter inkluderet i de primære analyser. Af alle intuberede patienter var 2,38% (696/29.209) uventet vanskelige at intubere på SARI afdelingerne mod 2,39% (723/30.305) på "Usual care" afdelingerne. Den justerede odds ratio var 1,03 (95% CI: 0,77-1,38), P = 0,84. Vi kunne ej heller

påvise signifikante forskelle på andre justerede effektmål og hverken en 58% stigning i antallet af patienter, der var forventet vanskelige at intubere eller en 84% stigning i patienter planlagt til avanceret intubationsteknik på SARI afdelingerne nåede et statistisk signifikant niveau, P = 0,29 og P = 0,06, henholdsvis.

Konklusion

Artiklerne der udgør denne ph.d. afhandling tydeliggør, at en høj andel af alle situationer med vanskelig luftvejshåndtering opstår uventet. I et cluster randomiseret forsøg var det ikke muligt at nedbringe andelen af uventet vanskelig intubation i en klinisk dagligdag gennem en systematisk tilgang til luftvejsvurdering i sammenligning med vanlig praksis. Indførelse af SARI modellen kan muligvis øge forventningen om intubationsvanskeligheder og kan muligvis påvirke klinisk praksis i retning af, at flere patienter bliver allokeret til intubation ved hjælp af avanceret udstyr. Denne afhandling understreger de fortsatte udfordringer, som anæstesilægen dagligt møder i bestræbelserne på at forudsige luftvejsrelaterede problemer.

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, sufficient ventilation and oxygenation is re-established by the anaesthetist. 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).

Pre-operative 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 pre-operative assessment and identification of patients at risk of airway difficulties [23]. It has been internationally accepted, that the pre-operative 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 pre-operative 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 pre-operative airway assessment recorded [24]. One of the main recommendations from the NAP4 was to perform thorough pre-operative 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 pre-operative airway assessment, it is still

unclear as to how this assessment should be performed. The American Society of Anesthesiologists (ASA) recommends a pre-operative 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 intu-

bation 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 pre-operative 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.

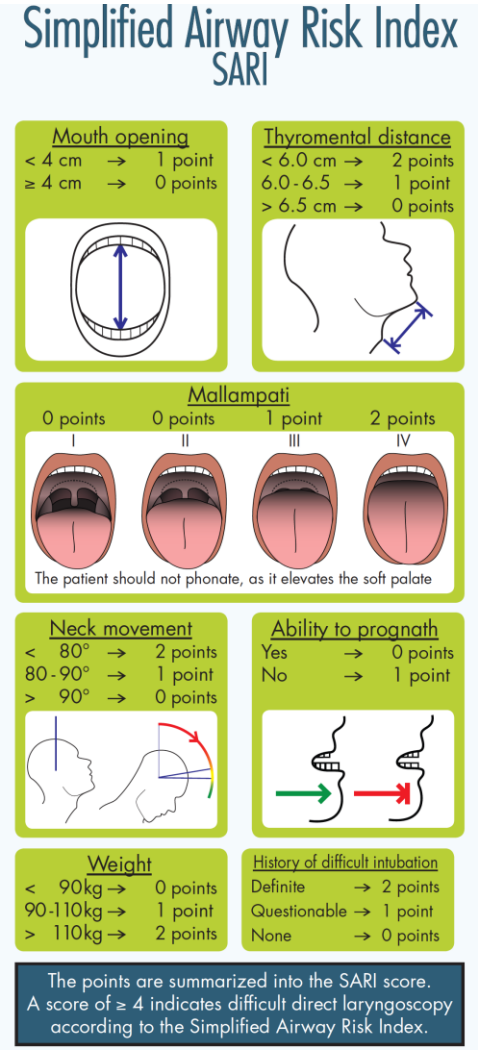
Predictive models

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 predic-

tive 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

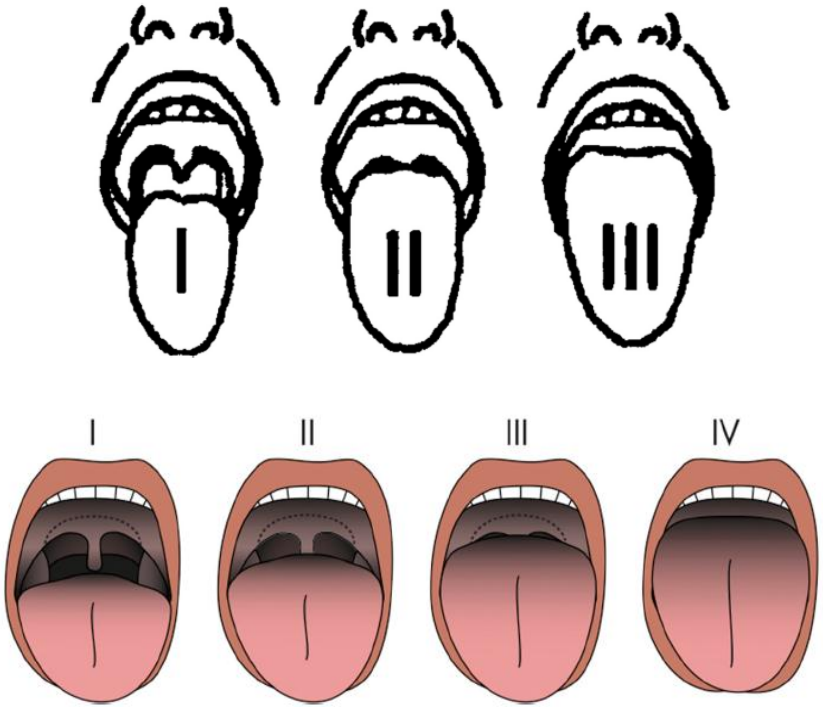


The SARI consists of 7 independent risk factors for difficult intubation.

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).

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



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.

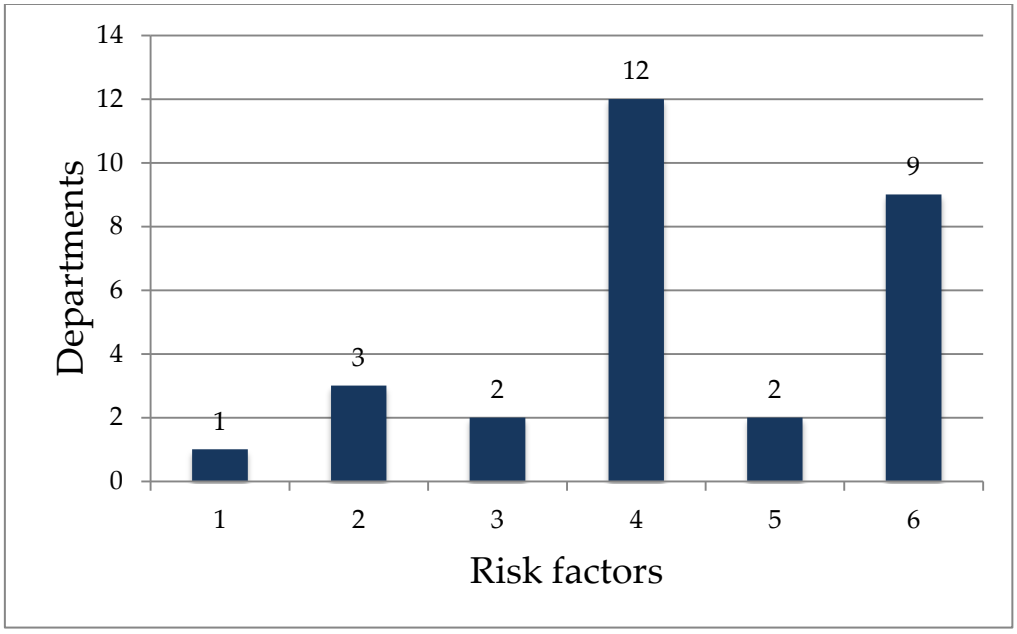
Pre-operative airway assessment in Denmark before initiation of this PhD

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 pre-operative airway examinations, depending on the individual anaesthesiologist and departmental recommendations.

Prior to engaging in a trial testing a predictive model it was important to establish whether pre-operative 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 pre-operative airway assessment, in some form, was widely practiced in Denmark, and we therefore found it fair to assume, that the anaes-

thesiologists pre-operative assessments were based on one, or several known risk factors of difficult airway management [13] (Figure 3).

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



Furthermore, there is reason to believe, that there may be a certain variance in the performance of the pre-operative 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.

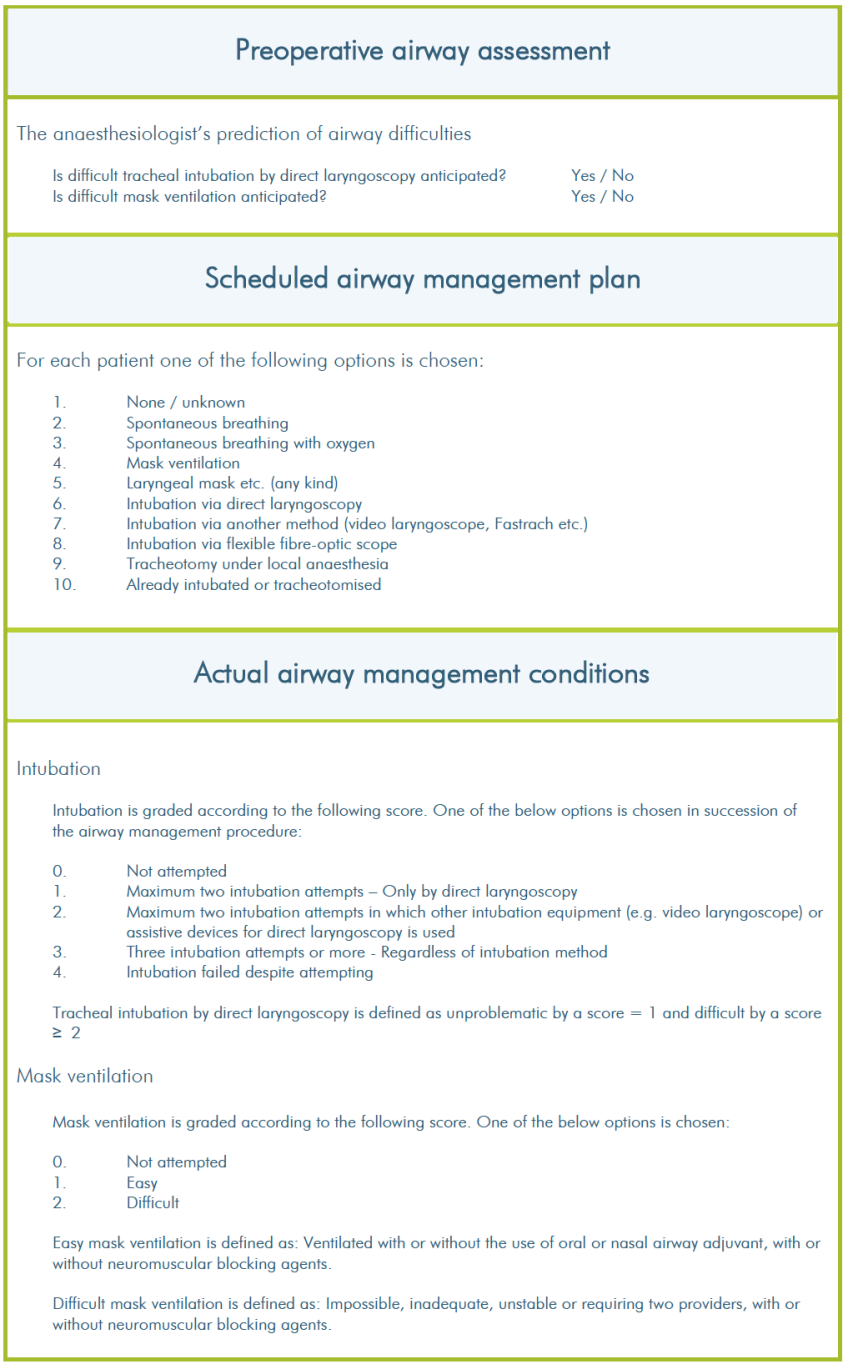
The Danish Anaesthesia Database

The Danish Anaesthesia Database is a national clinical quality assurance database containing selected quantifiable indicators, covering the anaesthetic process from the pre-operative assessment, through anaesthesia and surgery, until discharge from the post-anaesthesia 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 pre-operative 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. 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 re-

gistration. 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].

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



Preoperative airway assessment

The anaesthesiologist's prediction of airway difficulties

Is difficult tracheal intubation by direct laryngoscopy anticipated?	Yes / No
Is difficult mask ventilation anticipated?	Yes / No

Scheduled airway management plan

For each patient one of the following options is chosen:

1. None / unknown
2. Spontaneous breathing
3. Spontaneous breathing with oxygen
4. Mask ventilation
5. Laryngeal mask etc. (any kind)
6. Intubation via direct laryngoscopy
7. Intubation via another method (video laryngoscope, Fastrach etc.)
8. Intubation via flexible fibre-optic scope
9. Tracheotomy under local anaesthesia
10. Already intubated or tracheotomised

Actual airway management conditions

Intubation

Intubation is graded according to the following score. One of the below options is chosen in succession of the airway management procedure:

0. Not attempted
1. Maximum two intubation attempts – Only by direct laryngoscopy
2. Maximum two intubation attempts in which other intubation equipment (e.g. video laryngoscope) or assistive devices for direct laryngoscopy is used
3. Three intubation attempts or more - Regardless of intubation method
4. Intubation failed despite attempting

Tracheal intubation by direct laryngoscopy is defined as unproblematic by a score = 1 and difficult by a score ≥ 2

Mask ventilation

Mask ventilation is graded according to the following score. One of the below options is chosen:

0. Not attempted
1. Easy
2. Difficult

Easy mask ventilation is defined as: Ventilated with or without the use of oral or nasal airway adjuvant, with or without neuromuscular blocking agents.

Difficult mask ventilation is defined as: Impossible, inadequate, unstable or requiring two providers, with or without neuromuscular blocking agents.

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 pre-operative 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 pre-operative 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 pre-operative 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

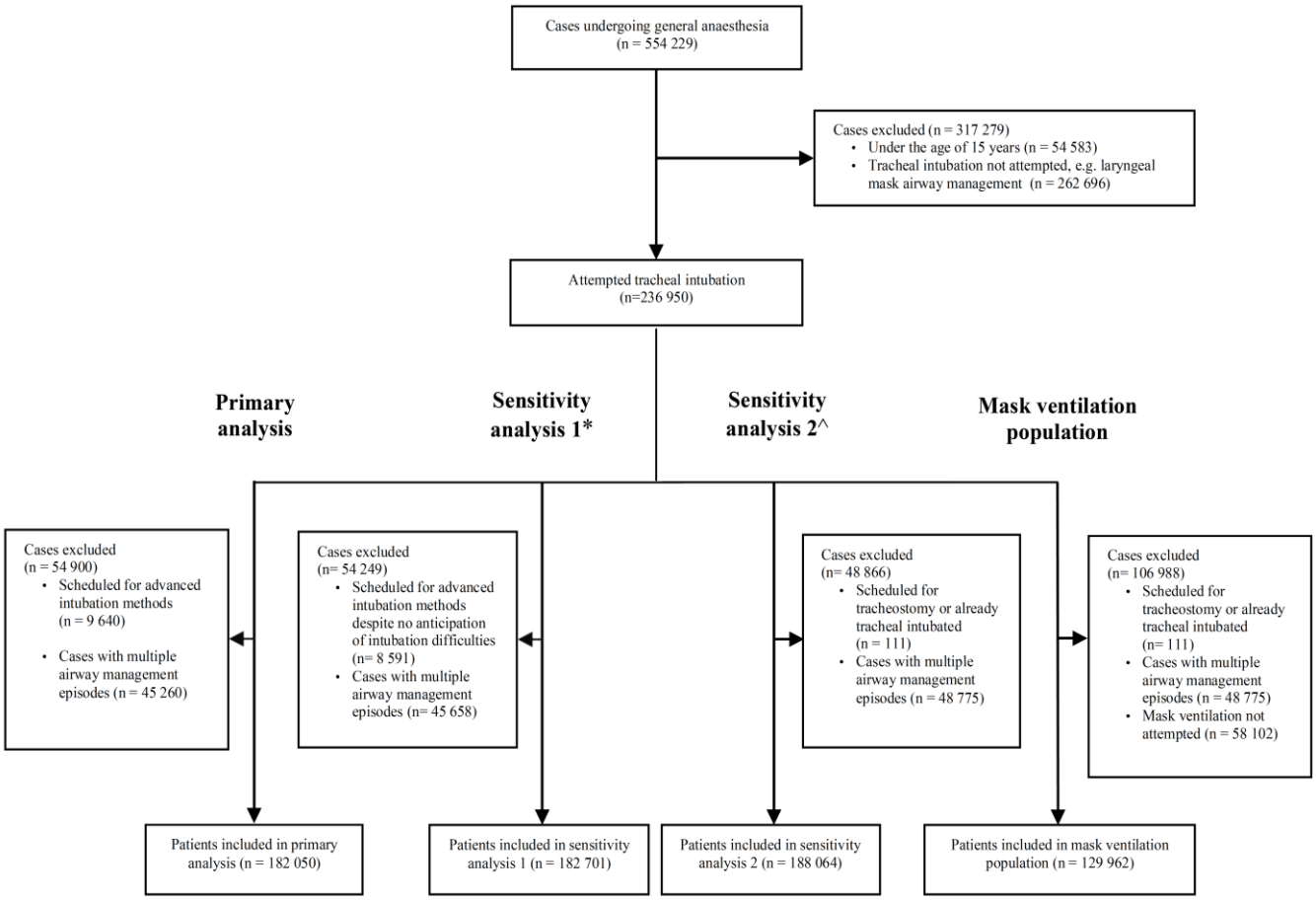
In the Danish Anaesthesia Database we identified 188,064 patients who had undergone tracheal intubation from June 1st 2008 to June 1st 2011 (Figure 5). The anaesthesiologists' pre-operative predictions on intubation and mask ventilation difficulties were compared with actual airway management conditions.

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%).

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.

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.

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 pre-operative 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 pre-operative 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 pre-operatively scheduled for advanced intubation techniques. Population 2: patients from Population 1 plus patients anticipated difficult to intubate, scheduled for, and attempted intubated by an advanced technique. Outcomes were assessed for both populations.

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

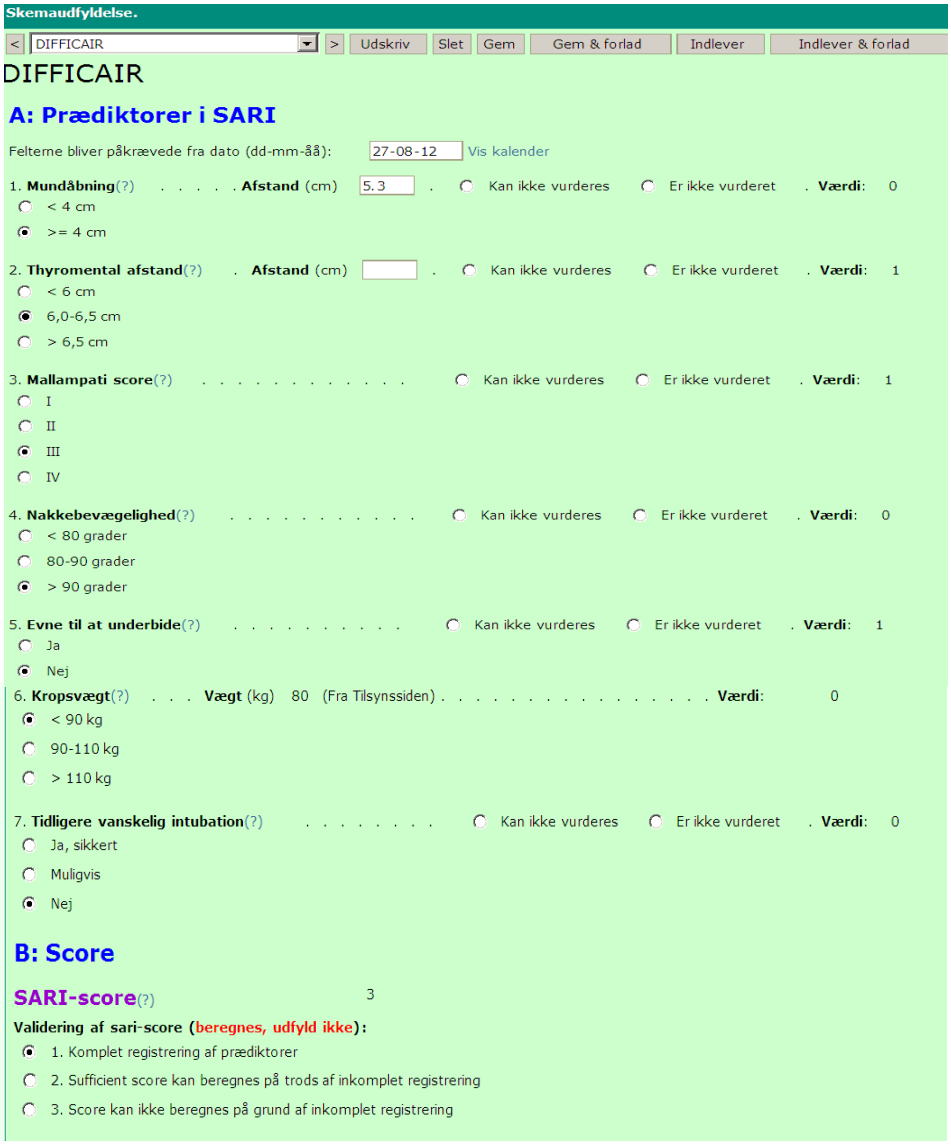
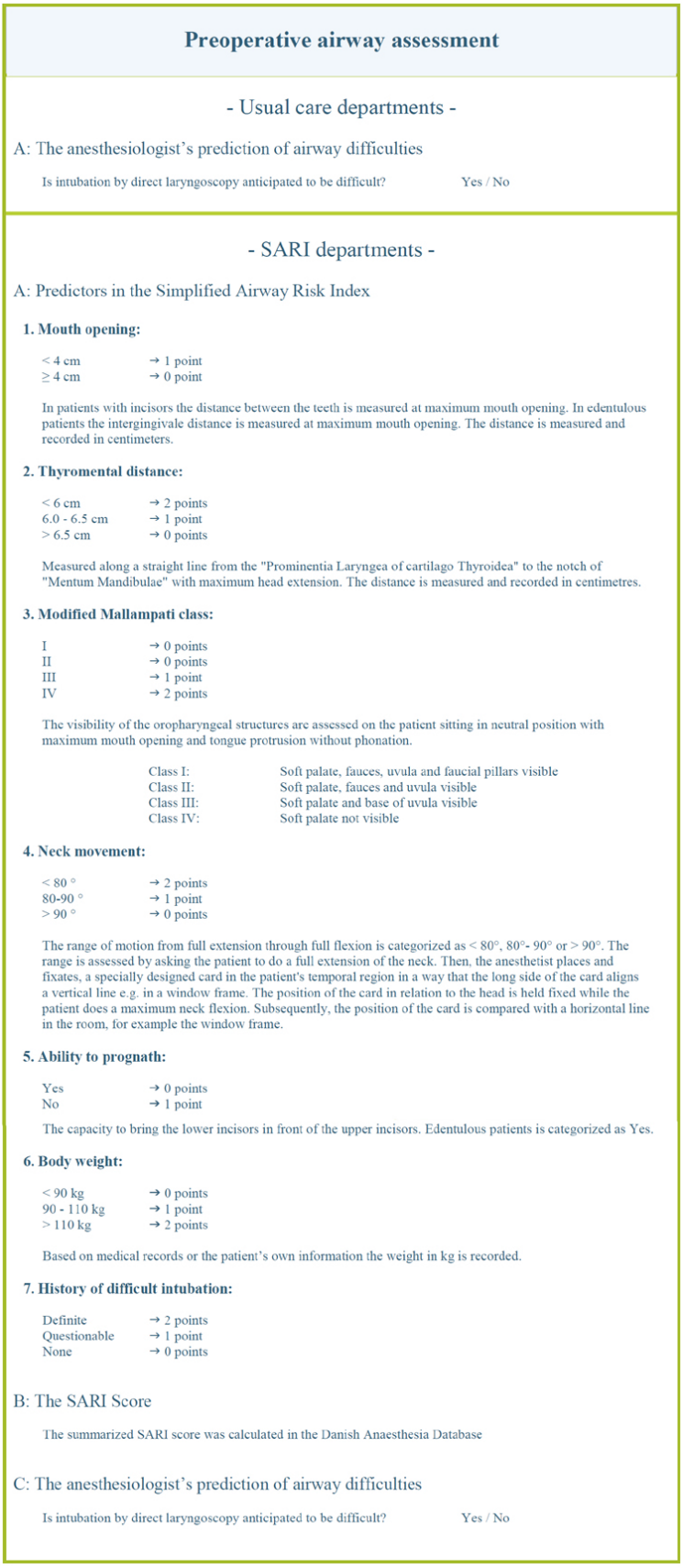


Figure 7. Pre-operative registration in the DAD



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

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.

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)).

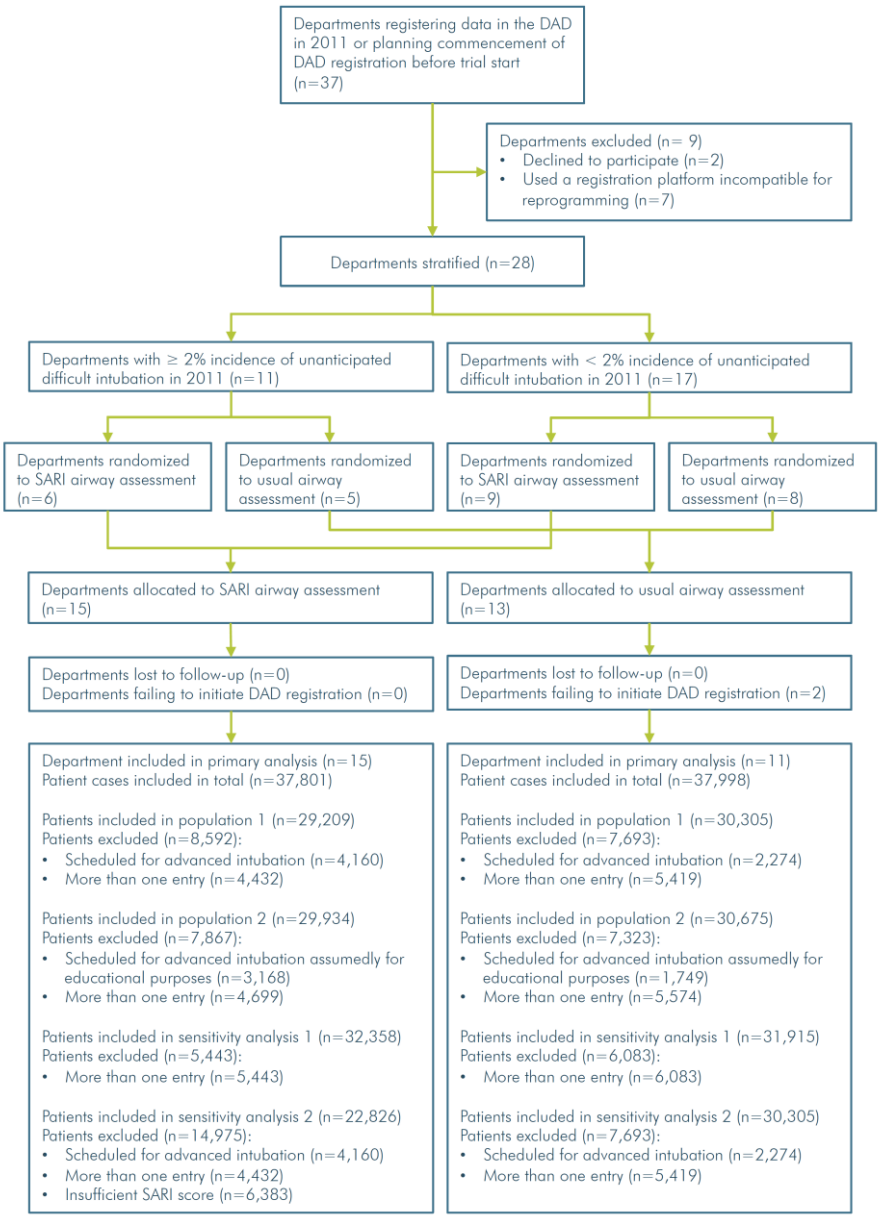
Primary outcomes

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.

Figure 8. Flow diagram of the DIFFICAIR trial



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.

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 standalone 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 prima-

ry 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 pre-specifying 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 pre-defined 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 sensitivi-

ty 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 pre-assumed 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 pre-operative 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 pre-operative 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 pre-operative 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 pre-operative 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 air-

way assessment, for example the number of pre-operative 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 pre-operative airway assessment compared to usual care based on the DIFFICAIR trial.

The SARI has now become recommended for pre-operative airway assessment in several departments; introduced in the chapter on pre-operative 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 pre-operative 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. Internal 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.

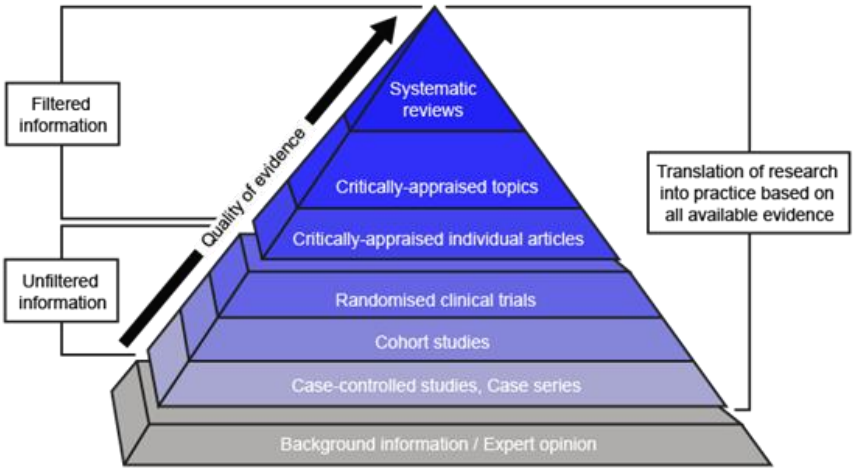
The cluster randomised trial

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).

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



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), 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 in-

creasing 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.

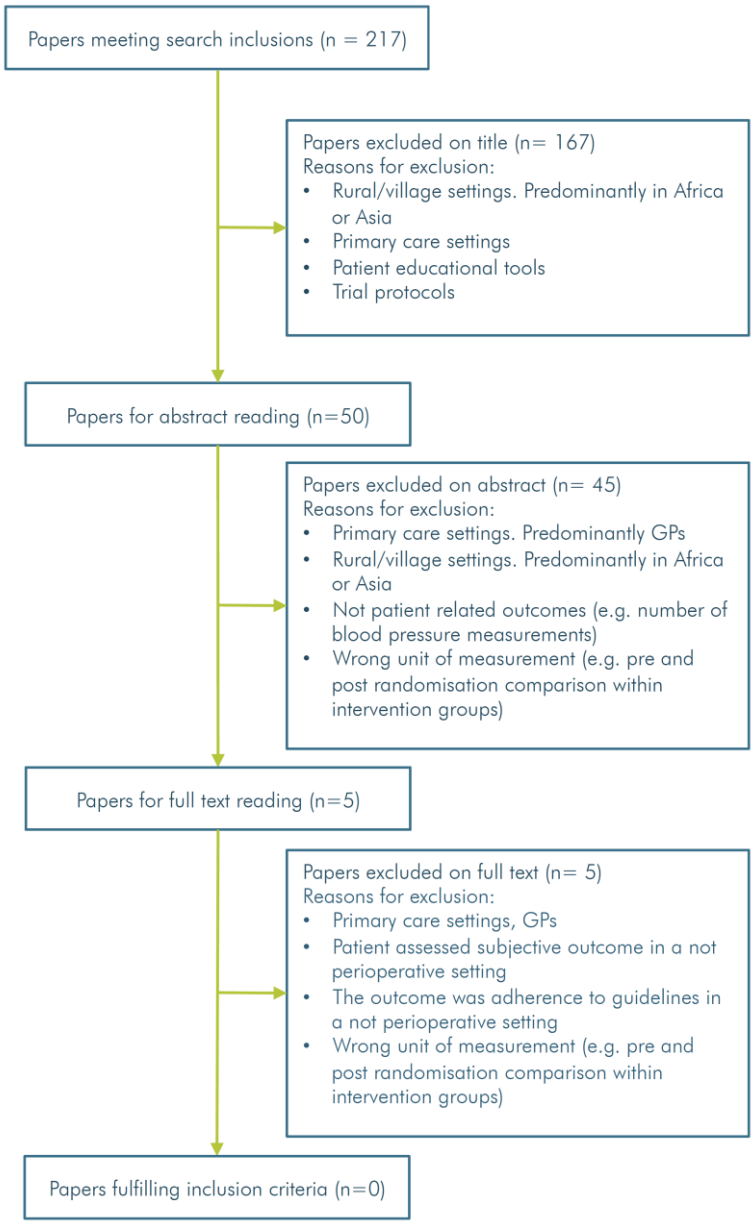
Methodology of the PhD in the context of medical research

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 peri-operative or in-hospital predictive or prognostic models. We found it interesting to elucidate, whether state of the art methodology for test-

ing the clinical impact of a prognostic model or recommendation had been applied before in a peri-operative 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 peri-operative 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 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 peri-operative setting.

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



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 peri-operative 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 evidence-based development and evolvement of clinical databases.

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 Unique™: 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.

Paper I - IV

Anaesthesia 2015, 70, 272–281 doi:10.1111/anae.12955

Original Article

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

A. K. Nørskov,[1,2] C. V. Rosenstock,[3] J. Wetterslev,[4] G. Astrup,[5] A. Afshari[6] and L. H. Lundstrøm[3]

1 Research Fellow, 3 Associate Professor, Department of Anaesthesiology, Copenhagen University Hospital, Nordsjællands Hospital, Hillerød, Denmark
2 Research Fellow, 4 Chief Physician, Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
5 Consultant, Department of Anaesthesiology, Aarhus University Hospital, Aarhus C, Denmark
6 Consultant, Department of Anaesthesiology, Juliane Marie Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark

Summary

Both the American Society of Anesthesiologists and the UK NAP4 project recommend that an unspecified pre-operative airway assessment be made. However, the choice of assessment is ultimately at the discretion of the individual anaesthesiologist. We retrieved a cohort of 188 064 cases from the Danish Anaesthesia Database, and investigated the diagnostic accuracy of the anaesthesiologists' predictions of difficult tracheal intubation and difficult mask ventilation. Of 3391 difficult intubations, 3154 (93%) were unanticipated. When difficult intubation was anticipated, 229 of 929 (25%) had an actual difficult intubation. Likewise, difficult mask ventilation was unanticipated in 808 of 857 (94%) cases, and when anticipated (218 cases), difficult mask ventilation actually occurred in 49 (22%) cases. We present a previously unpublished estimate of the accuracy of anaesthesiologists' prediction of airway management difficulties in daily routine practice. Prediction of airway difficulties remains a challenging task, and our results underline the importance of being constantly prepared for unexpected difficulties.

Correspondence to: A. Nørskov
Email: anderskehlet@hotmail.com
Accepted: 30 October 2014
*This paper is accompanied by an editorial: Anaesthesia 2015; **70**: 244-9.*

Introduction

Unanticipated difficult mask ventilation and difficult intubation may cause serious complications [1–4]. Accurate prediction of difficult airway management may reduce potential complications by allowing the allocation of experienced personnel and the use of relevant equipment [5]. No single predictor of difficult intubation is sufficiently reliable [6–11], and meta-analyses have found either none, or only sparse evidence, for a pre-operative assessment based on a single risk factor [6, 7, 11].

Difficult mask ventilation is associated with difficult intubation [12], and a situation with both difficult mask ventilation and difficult intubation is potentially life-threatening. Few studies have examined predictors for, as well as the proportions of, difficult mask ventilation

[13, 14]. The American Society of Anesthesiologists (ASA) recommends a pre-operative assessment of the patient's airway, based on eleven anatomical variables [15, 16], but without any elaboration 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. Consequently, it is left to the discretion of the individual anaesthesiologist [15, 16]. Likewise, in the UK, the recently published NAP4 project also recommends an unspecified pre-operative airway assessment [17].

Just as in the UK and USA, there are no specific national recommendations for pre-operative airway assessment in Denmark. Consequently, in daily routine practice, prediction of airway management difficulties is based on the individual anaesthesiologist's subjective answers to the following questions: (1) Do I anticipate difficult tracheal intubation? (2) Do I anticipate difficult mask ventilation? The answers to these questions may or may not be based on a diverse array of pre-operative airway examinations, depending on the individual anaesthesiologist and/or departmental recommendations.

Pre-operative airway assessment, in some form, is widely practised in Denmark, and it is likely that the subjective predictions are based on the examination of one or several known predictors of difficult airway management [18]. Previously published studies on this topic have focused on the predictive value of a single risk factor, or the value of combining several known risk factors into a multivariable predictive model. This study allows a novel estimate of the diagnostic accuracy of anaesthesiologists' subjective prediction of airway management difficulties, pragmatically reflecting daily clinical practice.

The Danish Anaesthesia Database (DAD), a national clinical quality assurance database, requires mandatory answers to questions (1) and (2) on anticipation of difficult tracheal intubation and mask ventilation, and is filled in pre-operatively by the anaesthesiologists. In addition, details regarding the patient's actual airway management conditions are registered, enabling comparisons between predictions and actual events. We hypothesised that the anaesthesiologists' subjective predictions would confirm prior predictive tools and risk factors, showing poor to moderate sensitivity and positive predictive value, and high specificity and negative predictive value [19].

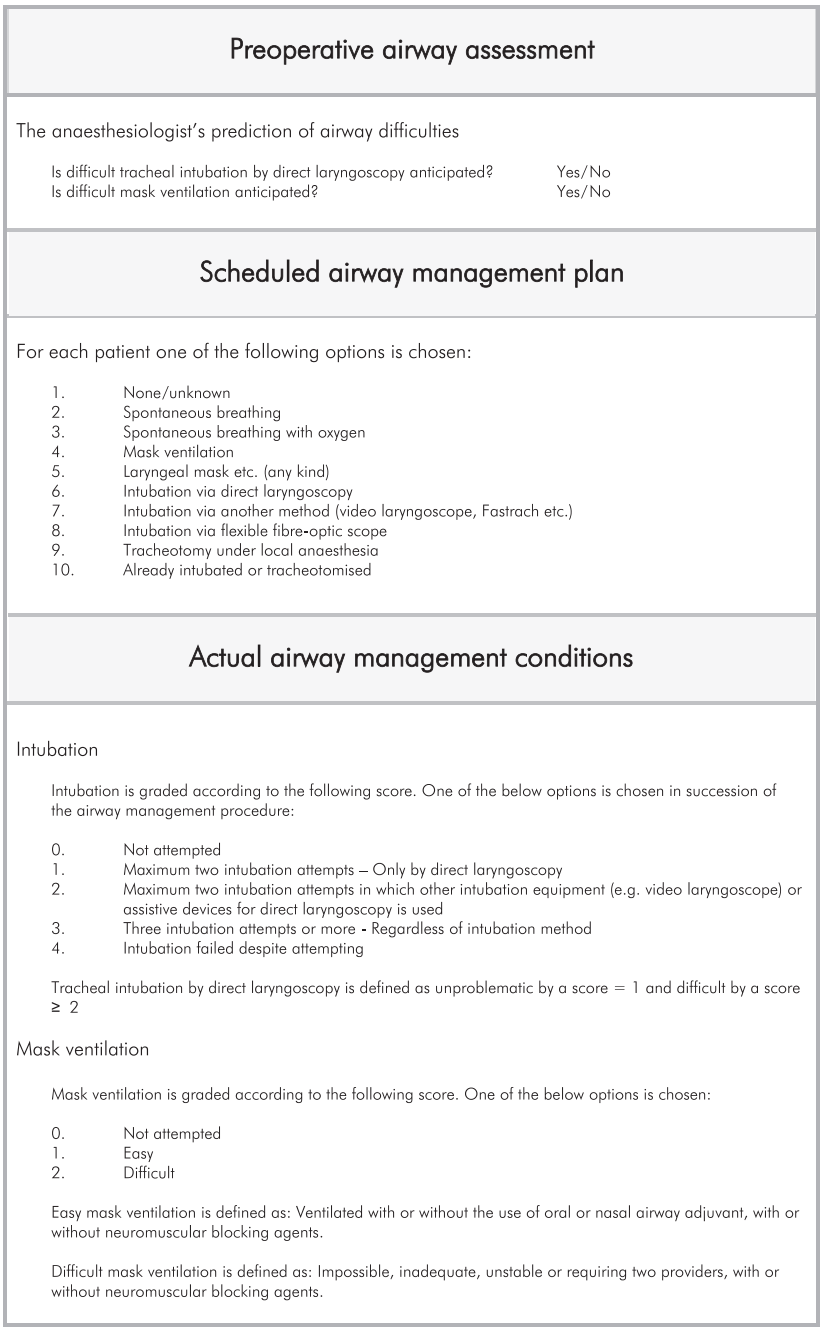
Using this pragmatic approach, which reflects the heterogeneous nature of everyday clinical practice, we aimed to estimate the accuracy with which subjective anticipation of airway difficulties predicts difficult tracheal intubation, and difficult mask ventilation.

Methods

This was an observational cohort study, and the Scientific Ethics Committee of Copenhagen County therefore waived the need for individual patient consent. Data extraction was approved by the Danish Data Protection Agency, and by the steering committee of DAD. In 2011, approximately 75% of all departments of anaesthesia in Denmark recorded data into the DAD. Prospective and consecutive data from all 37 Danish departments of anaesthesia recording in the DAD was extracted between 1 June 2008 and 1 June 2011.

The DAD is a national clinical quality assurance database containing selected quantifiable indicators, covering the anaesthetic process from the pre-operative assessment, through anaesthesia and surgery, until the end of the postoperative recovery period. Anaesthesiologists have to tick Y/N boxes to answer two mandatory questions regarding the anticipation of difficult tracheal intubation and difficult mask ventilation, following pre-operative airway assessment. In addition, a 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 on whom mask ventilation was attempted (Fig. 1). The National Board of Health and the Data Protection Agency approved the registration of all patients for anaesthesia.

We included all patients undergoing attempted intubation of their tracheas, and all patients in whom mask ventilation was attempted, in the study. The primary analysis was undertaken on patients who underwent attempted tracheal intubation initially with direct laryngoscopy. We then undertook two sensitivity analyses. The first (sensitivity analysis 1) included patients defined in the primary analysis, plus patients predicted by the anaesthesiologist as having tracheas difficult to intubate, who were scheduled for and under-



Preoperative airway assessment

The anaesthesiologist's prediction of airway difficulties

Is difficult tracheal intubation by direct laryngoscopy anticipated? Yes/No
Is difficult mask ventilation anticipated? Yes/No

Scheduled airway management plan

For each patient one of the following options is chosen:

1. None/unknown
2. Spontaneous breathing
3. Spontaneous breathing with oxygen
4. Mask ventilation
5. Laryngeal mask etc. (any kind)
6. Intubation via direct laryngoscopy
7. Intubation via another method (video laryngoscope, Fastrach etc.)
8. Intubation via flexible fibre-optic scope
9. Tracheotomy under local anaesthesia
10. Already intubated or tracheotomised

Actual airway management conditions

Intubation

Intubation is graded according to the following score. One of the below options is chosen in succession of the airway management procedure:

0. Not attempted
1. Maximum two intubation attempts – Only by direct laryngoscopy
2. Maximum two intubation attempts in which other intubation equipment (e.g. video laryngoscope) or assistive devices for direct laryngoscopy is used
3. Three intubation attempts or more - Regardless of intubation method
4. Intubation failed despite attempting

Tracheal intubation by direct laryngoscopy is defined as unproblematic by a score = 1 and difficult by a score ≥ 2

Mask ventilation

Mask ventilation is graded according to the following score. One of the below options is chosen:

0. Not attempted
1. Easy
2. Difficult

Easy mask ventilation is defined as: Ventilated with or without the use of oral or nasal airway adjuvant, with or without neuromuscular blocking agents.

Difficult mask ventilation is defined as: Impossible, inadequate, unstable or requiring two providers, with or without neuromuscular blocking agents.

Figure 1 Data registered in the Danish Anaesthesia Database.

went attempted tracheal intubation by an advanced intubation technique (e.g. videolaryngoscopy). The second (sensitivity analysis 2) was of all patients undergoing attempted tracheal intubation, regardless of technique. These are expanded upon below. In order to prevent bias, all patients were included with only one entry [20]. Patients having more than one episode of surgery in the recruitment period were included with their last data entry. We did not study children under the age of 15 years.

We investigated the following outcomes and analyses. Primary outcome measure: unanticipated difficult intubations [false negative]/all difficult intubations ([false negative] + [true positive]) =

1 − sensitivity. Secondary outcome measures included: (1) sensitivity = all correctly predicted difficult intubations [true positive]/all difficult intubations ([true positive] + [false negative]); (2) specificity = all correctly predicted easy intubations [true negative]/all easy intubations ([true negative] + [false positive]); (3) positive predictive value = all correctly predicted difficult intubations [true positive]/all intubations predicted difficult ([true positive] + [false positive]); (4) negative predictive value = all correctly predicted easy intubations [true negative]/all intubations predicted easy ([true negative] + [false negative]); (5) positive likelihood ratio = (sensitivity/(1 − specificity)); (6) negative likelihood ratio ((1 − sensitivity)/specificity); (7) odds ratio = ([true positive]/[false positive])/([false negative]/[true negative]).

We had two exploratory outcome measures: (i) the proportion of patients with anticipated difficult intubation scheduled for airway management by direct laryngoscopy; (ii) all-cause 30-day mortality in patients whose tracheas were intubated. Equivalent outcomes were measured for the mask ventilation population, and where appropriate, for the population of patients with both difficult tracheal intubation and difficult mask ventilation.

We used the definition of difficult intubation and difficult mask ventilation predefined in the DAD (Fig. 1). The intubation score was based on the number of intubation attempts, and the use of specialized airway equipment. The mask ventilation score was based on a simplified dichotomous version of Kheterpal and colleagues' definition of difficult mask ventilation [13, 21]. Patients whose tracheas were intubated by a more advanced technique than direct laryngoscopy were, according to the intubation score in DAD, categorised as having an airway that was difficult to intubate, regardless of the reason behind this choice (e.g. educational purposes). In the primary analysis we therefore excluded patients who were scheduled for intubation by advanced techniques before the operation. If a patient's trachea was subsequently intubated using advanced techniques, we assumed that the change of intubation equipment was due to difficulties with direct laryngoscopy, thus representing a difficult intubation (Fig. 2).

In order to avoid the erroneous exclusion of correctly identified difficult tracheal intubations, we conducted a first sensitivity analysis. We assumed that if the anaesthesiologist had predicted a difficult intubation, this was the reason for choosing an advanced technique. Consequently, we identified a group of patients who were predicted to have a difficult intubation, and who were scheduled for and underwent an advanced intubation technique. We assumed that these patients were correctly identified as difficult to intubate, and they were included as true positives, in addition to the patients in the primary analysis.

To explore if the accuracy of the predictions would improve when the severity of the intubation difficulties increased, we performed a second sensitivity analysis. We chose a more rigorous definition of difficult intubation than the predefined definition in DAD, thus defining difficult intubation as an intubation score ≥ 3. We included all patients who underwent attempted tracheal intubation.

As the study was an observational cohort study on a fixed available sample (DAD data 2008–2011), presenting frequencies of events, the number of patient cases entered into the DAD during the study period determined the sample size. The diagnostic accuracy of a subjective prediction of a difficult intubation, difficult mask ventilation, and the combination of both, was measured by: sensitivity; specificity; positive and negative predictive values; positive and negative likelihood ratios; and diagnostic odds ratio with 95% CI [22, 23]. A multivariable logistic regression analysis was used to adjust the odds ratios for potential confounders. The adjusted analysis included the following potential confounders: sex; age; body mass index; ASA physical status; use of neuromuscular blocking agents; and surgical priority (elective/emergency) [8, 9]. For all statistical analysis, we used SPSS v.22.0.0 (IBM Corp., Armonk, NY, USA). The variables used in this study were all mandatory for registration in the DAD, and thus we believe that the dataset was complete, with no missing variables. We reported the study according to the STROBE criteria [24].

Results

We found a total of 3383 (1.86%) difficult tracheal intubations registered. The number of unanticipated difficult intubations was 3154 (1.73%), and the primary outcome measure, the fraction of difficult intubations

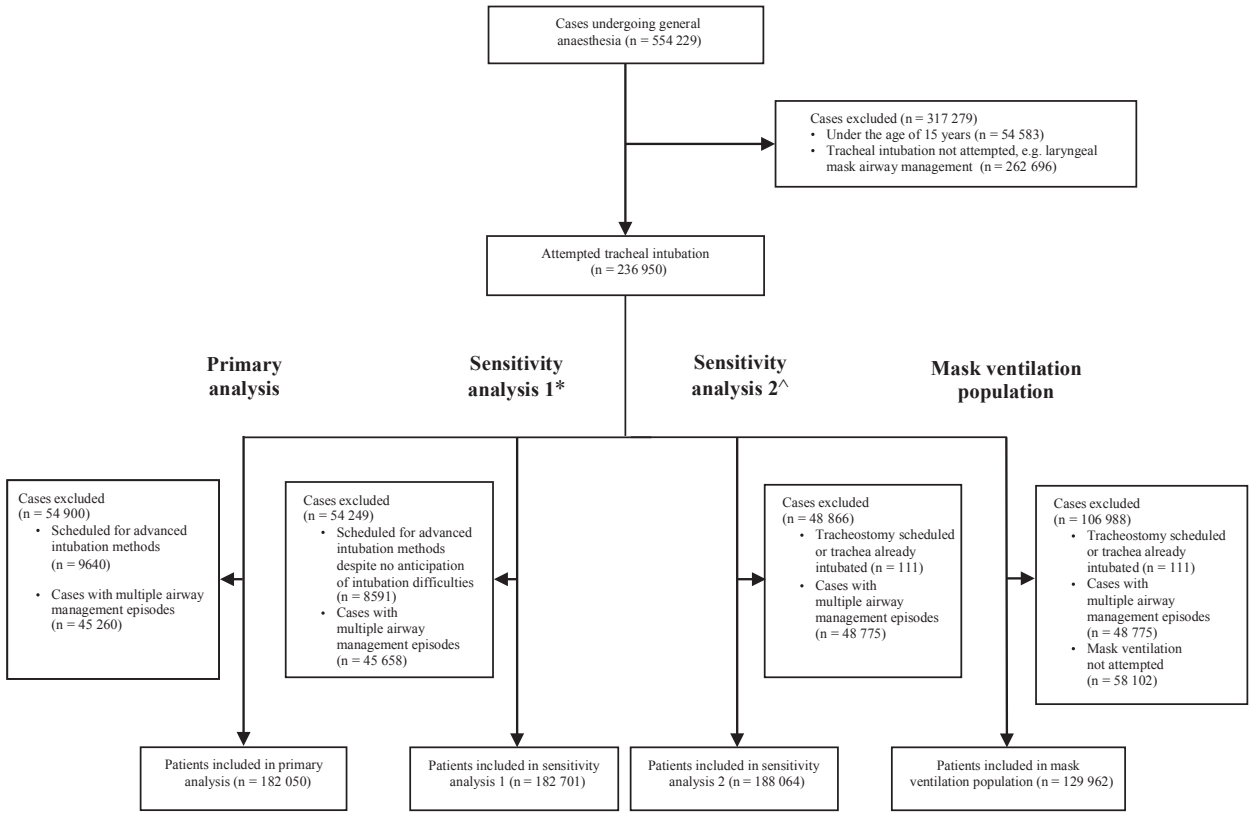


Figure 2 Flow diagram of the study populations. *First sensitivity analysis: this population includes, in addition to the population in the primary analysis, a group of patients predefined as having correctly identified difficult intubation, in whom intubation was anticipated to be difficult, and scheduled for and performed by more advanced methods than direct laryngoscopy (e.g. videolaryngoscopy). ^Second sensitivity analysis: this population includes all patients undergoing attempted tracheal intubation. Difficult intubation is defined as an intubation score ≥ 3.

that were not predicted pre-operatively, was 93% per cent (Table 1). In 25% of the cases where the anaesthesiologist anticipated difficult tracheal intubation, a difficult intubation actually occurred (the positive predictive value). The positive likelihood ratio was 17.28, reflecting how much the probability of difficult intubation increases if the patient tests positive (difficult intubation is anticipated). The adjusted odds ratio for a difficult intubation, when anticipated, was 17.04.

In sensitivity analysis 1, we categorised a group of patients as true positive cases. As a consequence, the number of correctly identified difficult tracheal intubation increased to 1060, corresponding to 75% of all difficult intubations' being unanticipated. The positive predictive value was 60%, and the positive likelihood ratio was 65.19. The adjusted odds ratio was 85.85 (Table 2).

In sensitivity analysis 2, tightening the definition of difficult intubation, there were 2657 difficult tracheal intubations. There were 2422 unanticipated difficult intubations, equalling 91% of all difficult intubations (Table 3).

The all-cause 30-day mortality for all patients whose tracheas were intubated was 3.1%.

We found 857 (0.66%) cases of difficult mask ventilation (Table 4). The number of cases of unanticipated difficult mask ventilation was 808 (0.62%), equivalent to 94% of all cases of difficult mask ventilation. The positive predictive value was 22%, and the positive likelihood ratio was 43.68. The adjusted odds ratio was 32.91 (Table 4). The all-cause 30-day mortality for mask-ventilated patients was 1.9%.

Of 857 patients with difficult mask ventilation, 424 (49.5%) patients were also registered with a

Table 1 Diagnostic accuracy of the anaesthesiologists' prediction of difficult intubation for patients undergoing attempted intubation with direct laryngoscopy initially (primary analysis). Values are number (proportion and/or 95% CI).

	Difficult intubation		
	Yes	No	Total
Anticipated difficult intubation			
Yes	229	700	929
No	3154	177 967	181 121
Total	3383	178 667	182 050
Difficult intubation	3383 (1.86% (1.80–1.92%))		
Unanticipated difficult intubation	3154 (1.73% (1.67–1.79%))		
Primary outcome			
Unanticipated difficult intubation (1 − sensitivity)	0.93 (0.92–0.94)		
Secondary outcomes			
Sensitivity	0.07 (0.06–0.08)		
Specificity	1.00 (1.00–1.00)		
Positive predictive value	0.25 (0.22–0.28)		
Negative predictive value	0.98 (0.98–0.98)		
Positive likelihood ratio	17.28 (14.94–19.98)		
Negative likelihood ratio	0.94 (0.93–0.94)		
Odds ratio	18.46 (15.84–21.52)		
Adjusted odds ratio	17.04 (14.59–19.96)		

difficult tracheal intubation. Thus, difficult mask ventilation was associated with difficult intubation, with an odds ratio of 45.77 (95% CI 39.83–52.61). In 29 of the 424 patients (6.8%), either difficult intubation or difficult mask ventilation was anticipated. In 16 cases (3.8%), both difficult intubation and difficult mask ventilation were correctly anticipated, illustrating that in 89.4% of the patients whose lungs proved difficult to ventilate, no airway management difficulties were predicted (Table 5). Failed tracheal intubation occurred in 3.7% of cases of difficult mask ventilation, compared with 0.1% where mask ventilation was not difficult. Thus, failed intubation was associated with difficult mask ventilation, with an OR of 38.48 (25.98–57.00).

We identified 1757 patients in whom difficult tracheal intubation was predicted. Of these patients, 47.5% were scheduled for tracheal intubation by an advanced method, and 48.4% were scheduled for intubation by direct laryngoscopy. A total of 4.1% were not scheduled for intubation, but subsequently had their tracheas intubated (Table 5).

Table 2 Diagnostic accuracy of the anaesthesiologists' prediction of difficult intubation for patients with anticipated difficult intubation, scheduled for and performed by an advanced method, categorised as true positives (sensitivity analysis 1). Values are number (proportion and/or 95% CI).

	Difficult intubation		
	Yes	No	Total
Anticipated difficult intubation			
Yes	1060*	697†	1757
No	3105†	177 839†	180 944
Total	4165	178 536	182 701
Difficult intubation	4165 (2.28% (2.21–2.35%))		
Unanticipated difficult intubation	3105 (1.70% (1.64–1.76%))		
Primary outcome			
Unanticipated difficult intubation (1 − sensitivity)	0.75 (0.73–0.76)		
Secondary outcomes			
Sensitivity	0.25 (0.24–0.27)		
Specificity	1.00 (1.00–1.00)		
Positive predictive value	0.60 (0.58–0.63)		
Negative predictive value	0.98 (0.98–0.98)		
Positive likelihood ratio	65.19 (59.55–71.37)		
Negative likelihood ratio	0.75 (0.74–0.76)		
Odds ratio	87.10 (78.66–96.45)		
Adjusted odds ratio	85.85 (77.23–95.42)		

*The 1060 patients include those in whom intubation was initially attempted by direct laryngoscopy and those in whom intubation was anticipated to be difficult, and planned for and intubated by an advanced method.
†The subtotals vary slightly from those in the primary analysis after including additional patients and checking that each patient was included only once in the analysis.

Discussion

This study presents a novel, and previously unpublished, estimate of the diagnostic accuracy of predictions of difficult airway management in daily clinical practice. The high positive likelihood ratio indicates that the anaesthesiologists' predictions of anticipated difficult airways are a strong diagnostic test. However, the high proportion of unanticipated difficult tracheal intubation, and the low positive predictive values, considerably reduces its value as a reliable diagnostic test in a clinical context.

Table 3 Diagnostic accuracy of the anaesthesiologists' prediction of difficult intubation for all patients undergoing attempted tracheal intubation; difficulty defined as an intubation score ≥ 3. Values are number (proportion and/or 95% CI).

	Difficult intubation		
	Yes	No	Total
Anticipated difficult intubation			
Yes	235	1487	1722
No	2422	183 920	186 342
Total	2657	185 407	188 064
Difficult intubation	2657 (1.41% (1.36–1.47%))		
Unanticipated difficult intubation	2422 (1.29% (1.24–1.34%))		
Primary outcome			
Unanticipated difficult intubation (1 − sensitivity)	0.91 (0.90–0.92)		
Secondary outcomes			
Sensitivity	0.09 (0.08–0.10)		
Specificity	0.99 (0.99–0.99)		
Positive predictive value	0.14 (0.12–0.15)		
Negative predictive value	0.99 (0.99–0.99)		
Positive likelihood ratio	11.03 (9.66–12.59)		
Negative likelihood ratio	0.92 (0.91–0.93)		
Odds ratio	12.00 (10.40–13.85)		
Adjusted odds ratio	11.27 (9.74–13.05)		

Table 4 Diagnostic accuracy of the anaesthesiologists' prediction of difficult mask ventilation. Values are number (proportion and/or 95% CI).

	Difficult mask ventilation		
	Yes	No	Total
Anticipated difficult mask ventilation			
Yes	49	169	218
No	808	128 936	129 744
Total	857	129 105	129 962
Difficult intubation	857 (0.66% (0.62–0.70%))		
Unanticipated difficult intubation	808 (0.62% (0.58–0.66%))		
Primary outcome			
Unanticipated difficult mask ventilation (1 − sensitivity)	0.94 (0.92–0.96)		
Secondary outcomes			
Sensitivity	0.06 (0.04–0.08)		
Specificity	1.00 (1.00–1.00)		
Positive predictive value	0.22 (0.17–0.29)		
Negative predictive value	0.99 (0.99–0.99)		
Positive likelihood ratio	43.68 (32.01–59.60)		
Negative likelihood ratio	0.94 (0.93–0.96)		
Odds ratio	46.27 (33.41–64.06)		
Adjusted odds ratio	32.91 (23.26–46.55)		

In the first sensitivity analysis, the diagnostic accuracy of the prediction increased noticeably, as a result of defining potentially true positive cases as such. When these cases were included, the number of difficult tracheal intubations that were not predicted pre-operatively reduced from 93% to approximately 75%. The 'true' accuracy of the anaesthesiologists' predictions of intubation difficulties probably lies somewhere between the predictive values found in the two populations, with a tendency in the primary analysis to underestimate the predictive power, and a tendency in the sensitivity analysis to overestimate it. The predictive accuracy did not improve when we employed a more rigorous definition of difficult intubation, as in the second sensitivity analysis. In situations comprising both difficult intubation and difficult mask ventilation, the predictive accuracy rose. Nevertheless, 89.4% of combined difficult intubation and difficult mask ventilation were unanticipated in the primary population.

Remarkably, almost half of the patients with anticipated tracheal intubation difficulties were still scheduled for direct laryngoscopy. Furthermore, when both difficult intubation and difficult mask ventilation were anticipated, 42.9% of patients were scheduled for airway management by direct laryngoscopy. Our findings further underline the clear association between difficult mask ventilation and difficult intubation reported in previous studies [14].

The group of patients whose tracheas were intubated had a higher 30-day mortality compared with the group of patients who underwent mask ventilation. Many patients were included in both groups, as they underwent both mask ventilation and intubation. However, this finding illustrates that patients whose tracheas were intubated, and particularly the ones not also mask ventilated (e.g. undergoing rapid sequence induction), were more likely to have a bad outcome, possibly due to more severe underlying co-morbidity that might have influenced their management and outcome.

The proportions of difficult tracheal intubation and difficult mask ventilation were low compared with previous studies [8, 11, 13, 25]. This may partly be due to the fact that there is no international consensus on how to define difficult intubation, and it is often

Table 5 Predicted and actual outcomes in airway management. Values are number (95% CI).

Combined intubation and mask ventilation	
All intubated and mask ventilated	129 962 (100%)
Difficult intubation and difficult mask ventilation	424 (0.33% (0.30–0.36%))
Failed intubation and difficult mask ventilation	32 (0.02% (0.02–0.03%))
Difficult intubation and difficult mask ventilation	424 (100%)
Anticipated one or both of difficult intubation and difficult mask ventilation	45 (10.61% (7.68–13.54%))
Anticipated both difficult intubation and difficult mask ventilation	16 (3.77% (1.96–5.59%))
Failed intubation and difficult mask ventilation	32 (100%)
Anticipated one or both of difficult intubation or difficult mask ventilation	6 (18.75% (5.23–32.27%))
Anticipated difficult airway management	
Anticipated difficult intubation	1757 (100%)
Scheduled for intubation by an advanced method	835 (47.52% 45.19–49.86%))
Scheduled for intubation by direct laryngoscopy	850 (48.38% (46.04–50.71%))
Not scheduled for intubation, but subsequently intubated	72 (4.10% (3.17–5.02%))
Anticipated difficult intubation and difficult mask ventilation	387 (100%)
Scheduled for intubation by an advanced method	200 (51.68% (46.70–56.66%))
Scheduled for intubation by direct laryngoscopy	162 (41.86% (36.95–46.78%))
Not scheduled for intubation, but subsequently intubated	25 (6.46% (4.01–8.91%))

equated with, and described by, the laryngoscopic view classified by Cormack and Lehane [26].

We assume that anaesthesiologists' subjective predictions were based on one or several known predictors of difficult intubation. However, the diagnostic accuracy of these predictions were poor, compared with reported studies on stand-alone tests, and with objective risk scores combining several predictors of difficult intubation [6, 11, 27]. Prior studies have been conducted under rigorous trial settings, and some even validated on the same study population, in contrast to our study, which reflects everyday clinical practice. Comparisons should therefore be made with caution.

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 possible scenario, where intubation and mask ventilation are impossible. The DAD only allows answers of 'Yes' or 'No' to the questions 'Anticipation of difficult intubation' and 'Anticipation of difficult mask ventilation'. It would have been preferable to have more differentiated information on the anticipated difficulties, to allow some insight into the airway planning made before induction of anaesthesia. This information gap may, to some extent, explain why we found that direct laryngoscopy was planned for such a high proportion of patients despite anticipation of difficulties.

This study has a number of possible limitations. The trial 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. The proportions of difficult intubation and difficult mask ventilation were very stable over the three-year period, but the data were not externally validated through other productivity logs or databases, and thus potentially unrecognised registration errors are possible. Most departments monitor the registrations of their patients, and do follow-up registrations on missing patients. However, we cannot rule out that some patients undergoing anaesthesia were never registered in the database, potentially resulting in an unknown number of missing data points.

Registrants in the DAD were unaware that the study was being conducted, and had no direct connection with the investigator group. However, each department providing data to the DAD could get access to pre-calculated data on the proportions of unanticipated intubation from their own department during the study period. The departments were not given specific data on the outcome, but data acquisition was in fact possible, if requested.

The conditions regarding tracheal intubation and mask ventilation were registered in the DAD following actual airway management. In case of an intubation score of ≥ 2 or difficult mask ventilation, it was man-

datory to fill out difficult airway management details. The database was programmed so that it was impossible to record difficult airway management parameters before other mandatory anaesthesia variables, including finishing time of the anaesthesia. Unfortunately, this programming was not ideal, and may have created an incentive to register intubation scores < 2, in order to avoid the extra registration burden. Ideally, registration of airway management difficulties ought to take place in real time, i.e. immediately following airway management. The airway manager would thus be encouraged to register correctly, instead of leaving the registration to be completed, potentially by another anaesthetist, at a later time. This may have led to underreporting of difficult airway management.

We found a surprisingly high percentage of unanticipated difficult airway management. This may partly be explained by the fact that the DAD registration frame was set by default to 'No' regarding anticipated difficult intubation and difficult mask ventilation. Though it is continuously reinforced in the departments that the anaesthesiologists should register correctly, as well as record their anticipation of a difficult airway management, we cannot exclude that the default settings may have affected the registrations, thereby creating a higher proportion of unanticipated difficult airway managements. Furthermore, it may be a misconception to assume that every practitioner performed an airway assessment, every time.

The reported proportion of unanticipated difficult tracheal intubation in daily routine practice, ranging from 75 to 93%, underlines the importance of always being prepared for unexpected airway management difficulties. Prediction of difficulties remains a challenging task. There may be ample room for improvement, based on a rigorous, evidence based and systematic approach [28].

Acknowledgements

We thank The Danish Anaesthesia Database and the participating departments.

Funding

The trial is funded by independent grants by: The TRYG foundation; Nordsjællands Hospitals research council; The Capital Regions funds for research; DASAIMs fund.

Competing interests

No competing interests.

References

1. Cooper GM, McClure JH. Anaesthesia chapter from Saving mothers' lives; reviewing maternal deaths to make pregnancy safer. *British Journal of Anaesthesia* 2008; **100**: 17–22.
2. Hove LD, Steinmetz J, Christoffersen JK, et al. 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.
3. Peterson GN, Domino KB, Caplan RA, Posner KL, Lee LA, Cheney FW. Management of the difficult airway – a closed claims analysis. *Anesthesiology* 2005; **103**: 33–9.
4. Rosenstock C, Møller J, Hauberg A. Complaints related to respiratory events in anaesthesia and intensive care medicine from 1994 to 1998 in Denmark. *Acta Anaesthesiologica Scandinavica* 2001; **45**: 53–8.
5. Cook TM, Macdougall-Davis SR. Complications and failure of airway management. *British Journal of Anaesthesia* 2012; **109**(Suppl 1): i68–85.
6. 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. *British Journal of Anaesthesia* 2011; **107**: 659–67.
7. 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. *Anesthesia and Analgesia* 2006; **102**:1867–78.
8. 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**: 266–74.
9. Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Gätke MR, Wetterslev J. Avoidance of neuromuscular blocking agents may increase the risk of difficult tracheal intubation: a cohort study of 103,812 consecutive adult patients recorded in the Danish Anaesthesia Database. *British Journal of Anaesthesia* 2009; **103**: 283–90.
10. 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.
11. Shiga T, Wajima Z, Inoue T, Sakamoto S, Sakamoto A. Predicting difficult intubation in apparently normal patients. *Anesthesiology* 2005; **103**: 429–37.
12. Kheterpal S, Healy D, Aziz MF, et al. 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.
13. Kheterpal S, Han R, Tremper KK, et al. Incidence and predictors of difficult and impossible mask ventilation. *Anesthesiology* 2006; **105**: 885–91.

14. Kheterpal S, Martin L, Shanks AM, Tremper KK. Prediction and outcomes of impossible mask ventilation: a review of 50,000 anesthetics. *Anesthesiology* 2009; **110**: 891–7.
15. 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**: 1269–77.
16. Apfelbaum JL, Hagberg CA, Caplan RA, et al. 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.
17. 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. *British Journal of Anaesthesia* 2011; **106**: 266–71.
18. Nørskov AK, Rosenstock CV, Wetterslev J, Lundstrøm LH. Preoperative airway assessment. No national consensus in Denmark. *European Journal of Anaesthesiology* 2014; **31**(e-suppl 52):275.
19. Yentis SM. Predicting difficult intubation – worthwhile exercise or pointless ritual? *Anaesthesia* 2002; **57**: 105–9.
20. Altman DG, Bland JM. Statistics notes. Units of analysis. *British Medical Journal* 1997; **314**: 1874.
21. Han R, Tremper K, Kheterpal S, O'Reilly M. Grading scale for mask ventilation. *Anesthesiology* 2004; **101**: 267.
22. Altman DG, Bland JM. Diagnostic tests. 1: sensitivity and specificity. *British Medical Journal* 1994; **308**: 1552.
23. Altman DG, Bland JM. Diagnostic tests 2: predictive values. *British Medical Journal* 1994; **309**: 102.
24. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Journal of Clinical Epidemiology* 2008; **61**: 344–9.
25. Randell T. Prediction of difficult intubation. *Acta Anaesthesiologica Scandinavica* 1996; **40**: 1016–23.
26. Cormack RS, Lehane J. Difficult tracheal intubation in obstetrics. *Anaesthesia* 1984; **39**: 1105–11.
27. El-Ganzouri AR, McCarthy RJ, Tuman KJ, Tanck EN, Ivankovich AD. Preoperative airway assessment: predictive value of a multivariate risk index. *Anesthesia and Analgesia* 1996; **82**: 1197–204.
28. 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.

TRIALS

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

Nørskov *et al.*

BioMed Central

Nørskov *et al. Trials* 2013, **14**:347
http://www.trialsjournal.com/content/14/1/347

TRIALS

STUDY PROTOCOL Open Access

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

Anders Kehlet Nørskov[1,2*], Charlotte Valentin Rosenstock[1], Jørn Wetterslev[2] and Lars Hyldborg Lundstrøm[1]

Abstract

Background: Pre-operative airway assessment in Denmark is based on a non-specific clinical assessment. Systematic, evidence-based and consistent airway assessment may reduce the incidence of unanticipated difficult airway management. By assessing multiple predictors for difficult airway management, the predictive value of the assessment increases. The Simplified Airway Risk Index (SARI) is a multivariate risk score for predicting difficult intubation.
This study aims to compare the use of the SARI with a non-specified clinical airway assessment on predicting difficult intubation. Further, to compare the examination and registration of predictors for difficult mask ventilation with a non-specified clinical airway assessment on prediction of difficult mask ventilation.

Method/Design: We cluster-randomized 28 Danish departments of anaesthesia to airway assessment either by the SARI or by usual non-specific assessment. Data from patients' pre-operative airway assessment are registered in the Danish Anaesthesia Database. Objective scores for intubation and mask ventilation grade the severity of airway managements. The accuracy of predicting difficult intubation and mask ventilation is measured for each group. The primary outcome measure is the fraction of unanticipated difficult and easy intubation.
The fraction of unanticipated difficult intubation in Denmark is 1.87%. With a stratified randomization, type 1 error risk of 5% and a power of 80%, 30 departments are required to detect or reject a 30% relative risk reduction equalling a number needed to treat of 180. Sample size estimation is adjusted for the study design and based on standards for randomization on cluster-level. With an average cluster size of 2,500 patients, 70,000 patients will be enrolled over a 1-year trial period. The database is programmed so that registration of the SARI and predictors for difficult mask ventilation are mandatory for the intervention group but invisible to controls.

Discussion: It is innovative to use a national clinical database as the basis for a randomized clinical trial. The method can serve as a precedent for implementation of evidence-based recommendations and database registration.
The trial will forward understanding of how to predict and reduce unanticipated difficult airways and how to produce evidence-based recommendations for airway assessment and clinical database development.

Trial registration: (NCT01718561).

Keywords: Airway management, Cluster analysis, Difficult intubation, Randomized controlled trials, Sensitivity, Specificity

* Correspondence: anderskehlet@hotmail.com
[1]Department of Anaesthesiology, Copenhagen University Hospital, Nordsjælland Hospital, 3400, Hillerød, Denmark
[2]Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen University Hospital, Rigshospitalet 2100, Copenhagen, Denmark

Background

Unanticipated difficult airways are dreaded amongst anaesthesiologists and difficult tracheal intubation and difficult mask ventilation (DMV) can cause serious patient complications [1-4]. Better prediction of unanticipated difficult airways may reduce morbidity and mortality by allocating experienced personnel and relevant equipment. There is no single predictor that is sufficiently valid in predicting difficult tracheal intubation [5-10]. However, several studies show that by combining multiple predictors of difficult tracheal intubation, the positive and the negative predictive value of the assessment increases [10].

Mask ventilation is an essential component of airway management during general anaesthesia. In the event of failed intubation, establishing successful mask ventilation and thus oxygenation of the patient can be a life-saving procedure; DMV is correlated with difficult tracheal intubation [11,12]. A situation with both DMV and difficult tracheal intubation may place the patient at serious risk of complications or even death. Few studies have examined predictors for DMV and the frequency of the event [11,12]. There are no clear recommendations for when a patient should be considered at risk of DMV.

The American Society of Anesthesiologists (ASA) recommends a preoperative assessment of the patient's airway based on 11 anatomical parameters [13,14]. Despite the ASA recommendation, there is no defined recommendation on which factors are mandatory for examination, nor on how these should be weighted in an overall airway assessment, and some of the critical cut-off values for the factors are not clearly defined. The ASA argues that the decision to assess some or all risk factors depends on the clinical context and it is left to the discretion of the individual anaesthesiologist [13,14]. In the UK, the Difficult Airway Society guidelines for management of the unanticipated difficult intubation [15] does not recommend preoperative airway assessment because of disputes about its value. However, the recently published Fourth National Audit Project [16] is opening up for a recommendation of a preoperative airway assessment although it has not been further defined.

The Danish Anaesthesia Database (DAD) is a clinical database that contains selected quantifiable indicators, covering the anaesthetic process from the preoperative assessment through anaesthesia and surgery to the post-operative recovery period. At present, all patient records in the database include the anaesthesiologist's unspecified assessment of potential airway difficulties as well as a scheduled airway management plan. For all patients receiving general anaesthesia with an attempted intubation, an airway management score is registered based on the conditions of the (attempted) intubation. Likewise, in patients with attempted mask ventilation, an airway management score is registered for (attempted) mask ventilation.

In agreement with the ASA recommendations, the preoperative airway assessment in DAD is currently based exclusively on the individual anaesthesiologist's preoperative clinical assessment, which is more or less based on various known, unknown, or less verified predictors of a difficult airway. Based on this assessment, whether mask ventilation and/or tracheal intubation by direct laryngoscopy is expected to be difficult is recorded as yes or no. Subsequently, the strategy for airway management is planned and recorded. There is little documentation of how accurately this preoperative clinical assessment predicts actual airway management conditions.

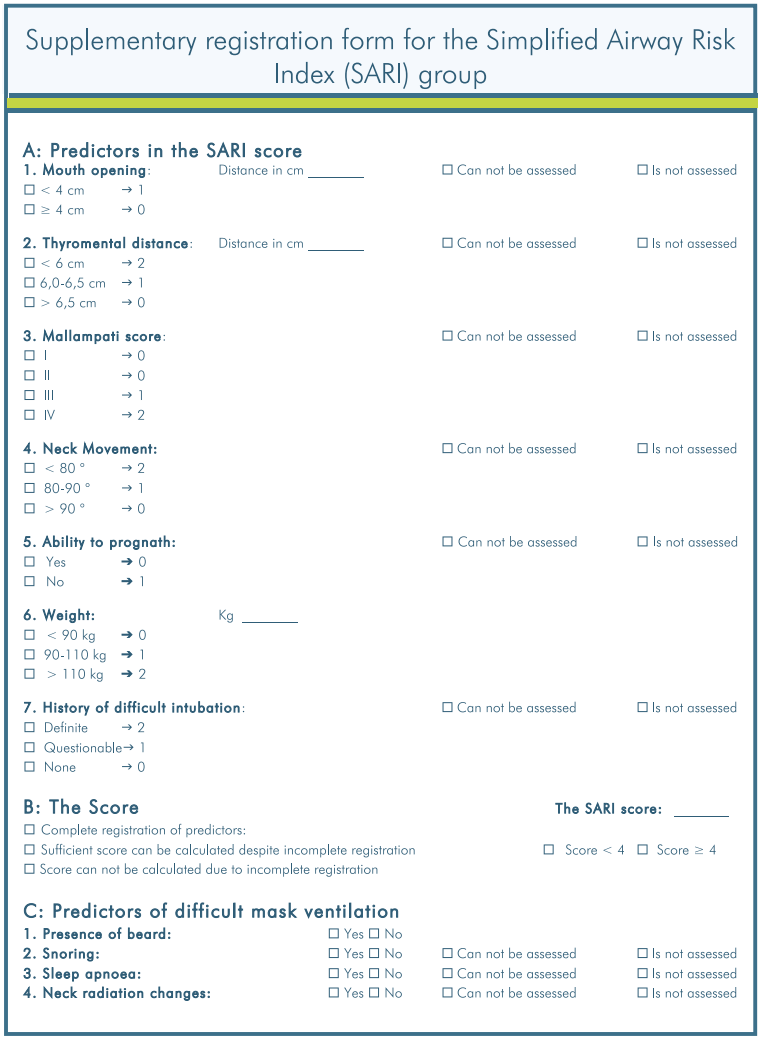
The Simplified Airway Risk Index (SARI) is a multivariate model for airway assessment described by El-Ganzouri et al., enabling an estimation of the likelihood of a difficult direct laryngoscopy [17]. The SARI contains seven individual predictors for difficult direct laryngoscopy, each given a weighted score 0–1 or 0–2. A summed value of the SARI score >3 indicates a future direct laryngoscopy to become difficult (Figure 1). It is unknown whether the SARI score predicts difficult intubation better or worse than a clinical assessment. We will compare the effect of using the SARI with an unspecified clinical airway assessment on the prediction of difficult intubation by direct laryngoscopy in a randomized clinical trial. Further, we want to record known risk factors for DMV and to investigate whether systematic registration of these risk factors leads to a reduction in DMV. During the DIFFICAIR trial, an internet page in the DAD will enable pre-operative registration of risk factors comprised in the SARI model. Kheterpal et al. described several risk factors associated with DMV [11,12]. Predictors for DMV will be a part of the data assessed and recorded in DAD in addition to the SARI score.

Null hypothesis

- There is no difference in the proportion of unanticipated difficult intubations when the preoperative airway assessment is based on the SARI score compared with a preoperative airway assessment based on the individual anaesthesiologist's assessment.
- There is no difference in the proportion of unanticipated DMV when the preoperative airway assessment includes systematic examination and registration of known predictors for DMV compared with an unstructured examination.

Methods/Design

The trial is a cluster (cluster = department) and parallel group randomized trial stratified for the proportion of unanticipated difficult intubation. A total of 28 Danish departments of anaesthesia participate in the



Supplementary registration form for the Simplified Airway Risk Index (SARI) group

A: Predictors in the SARI score

1. Mouth opening: Distance in cm _____ ☐ Can not be assessed ☐ Is not assessed
☐ < 4 cm → 1
☐ ≥ 4 cm → 0

2. Thyromental distance: Distance in cm _____ ☐ Can not be assessed ☐ Is not assessed
☐ < 6 cm → 2
☐ 6,0-6,5 cm → 1
☐ > 6,5 cm → 0

3. Mallampati score: ☐ Can not be assessed ☐ Is not assessed
☐ I → 0
☐ II → 0
☐ III → 1
☐ IV → 2

4. Neck Movement: ☐ Can not be assessed ☐ Is not assessed
☐ < 80 ° → 2
☐ 80-90 ° → 1
☐ > 90 ° → 0

5. Ability to prognath: ☐ Can not be assessed ☐ Is not assessed
☐ Yes → 0
☐ No → 1

6. Weight: Kg _____
☐ < 90 kg → 0
☐ 90-110 kg → 1
☐ > 110 kg → 2

7. History of difficult intubation: ☐ Can not be assessed ☐ Is not assessed
☐ Definite → 2
☐ Questionable → 1
☐ None → 0

B: The Score

The SARI score: _____

☐ Complete registration of predictors:
☐ Sufficient score can be calculated despite incomplete registration ☐ Score < 4 ☐ Score ≥ 4
☐ Score can not be calculated due to incomplete registration

C: Predictors of difficult mask ventilation

1. Presence of beard: ☐ Yes ☐ No
2. Snoring: ☐ Yes ☐ No ☐ Can not be assessed ☐ Is not assessed
3. Sleep apnoea: ☐ Yes ☐ No ☐ Can not be assessed ☐ Is not assessed
4. Neck radiation changes: ☐ Yes ☐ No ☐ Can not be assessed ☐ Is not assessed

Figure 1 Supplementary registration form for the SARI group. This form, or a similar sticker, is attached to the anaesthesia record in the SARI group.

DIFFICAIR trial. They are randomized 1:1 as intervention departments with systematic airway assessment according to the SARI score and registration in the DAD or as control departments with preoperative airway assessment based on the individual anaesthesiologists' assessment.

Randomization

We conducted a baseline study in 2011 using data from the DAD version 3 and determined the proportion of unanticipated difficult intubation for each department of anaesthesia. The departments were then stratified according to whether the proportion of unanticipated difficult intubation was above or below 2%.

With appropriate use of allocation concealment, the heads of departments provided written informed consent before the departments were randomized. Thereafter, according to a computer-generated list of the allocation sequence, the departments were randomly assigned to one of two groups. In one group, anaesthesio-logists are trained in preoperative use of the SARI score (the SARI group) and in a control group the preoperative airway assessments of the anaesthesiologists are based solely on a clinical assessment (CA group). The SARI group is thus included in a trial branch in which each patient has a preoperative airway assessment and a matching DAD registration consisting of a fixed panel of predetermined predictors for difficult intubation. In the SARI group four additional variables, which may be associated with DMV, are also recorded in DAD. Departments in the CA group continue to use an individual assessment of each patient regarding on whether the airway management will become difficult or not; this is preoperatively registered in DAD.

Cluster randomization vs. individual randomization

Anaesthesiologists taught the use of the SARI score will inevitably and unintentionally use this knowledge during airway assessments also when assessing patients randomized for a "clinical assessment" only [18,19]. Therefore, a trial design using individual randomization of anaesthesiologists and patients is prone to yield incorrect results for the comparison of the two assessment methods. This is due to a "spill-over" effect between the trial branches within departments. Accordingly, we chose to randomize patients clustered on a departmental level [19].

Inclusion

Departments registering patients in the DAD with an expected minimum of 200 intubations annually are eligible for inclusion.

Three populations of randomized patients are identified: Population 1: All patients primarily (attempted) intubated by direct laryngoscopy; Population 2: All patients primarily (attempted) intubated by direct laryngoscopy plus patients that are expected to be difficult to intubate by direct laryngoscopy and are therefore scheduled for intubation with an advanced method (e.g., video laryngoscopic or fibre-optic intubation); Population 3: All patients undergoing mask ventilation.

Exclusion

Children <15 years old.

Primary outcome measures

The following are measured regardless of randomization: i)Fraction of unanticipated difficult intubations = intubations with unanticipated difficulties [False negative]/all patients primarily (attempted) intubated by direct laryngoscopy; ii) Fraction of unanticipated easy intubations = intubations with anticipated difficulties that were easy [False Positive]/all patients primarily (attempted) intubated by direct laryngoscopy. Simultaneous low fractions of the primary outcome measures are desirable for good prediction of difficult intubation.

Secondary outcome measures

- 48-hour mortality
- 30-day mortality
- Fraction = intubations anticipated to be difficult, thus planned for, and intubated by, an advanced method/all patients (attempted) intubated
- Fraction = unanticipated difficult intubations [False Negative]/true difficult intubations ([False negative] + [True Positive])
- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value
- Positive Likelihood Ratio = (Sensitivity/(1-Specificity))
- Negative Likelihood Ratio = ((1-Sensitivity)/Specificity)
- The Receiver Operating Characteristic curve. A graphical representation of sensitivity as a function of (1-Specificity). Applicable for comparison of predictive models.

An analogue outcome measurement will be done for mask ventilation.

The simplified airway risk index (SARI)

The SARI model consists of seven predictors for difficult direct laryngoscopy:

1. Mouth opening
2. Thyromental distance
3. Mallampati class
4. Neck movement
5. Ability to prognath
6. Weight
7. History of difficult intubation

The SARI uses the original Mallampati grade, whereas for data entry in DAD a modified Mallampati class [20] will be used (Figure 2 and Figure 3). The Mallampati grades contribute to the SARI score as follows: Grade III → 2 points; Grade II → 1 point; Grade I → 0 points.

The original Mallampati grade I approximately corresponds to the modified Mallampati classes I and II, the original Mallampati grade II approximately corresponds to the modified Mallampati class III, and the original Mallampati grade III corresponds to the modified Mallampati class IV (Figure 4).

Predictors of difficult and impossible mask ventilation

The following parameters that correlate to difficult/impossible mask ventilation are registered in the SARI group:

1. Changes in the neck due to radiation
2. Presence of beard
3. BMI ≥30 kg/m^2
4. Age ≥57 years
5. Modified Mallampati score III or IV
6. Severely limited jaw protrusion
7. Snoring
8. Sleep apnoea

The predictors for DMV are already recorded in the DAD and the SARI except for the four listed below. Consequently, departments allocated to the SARI group also record:

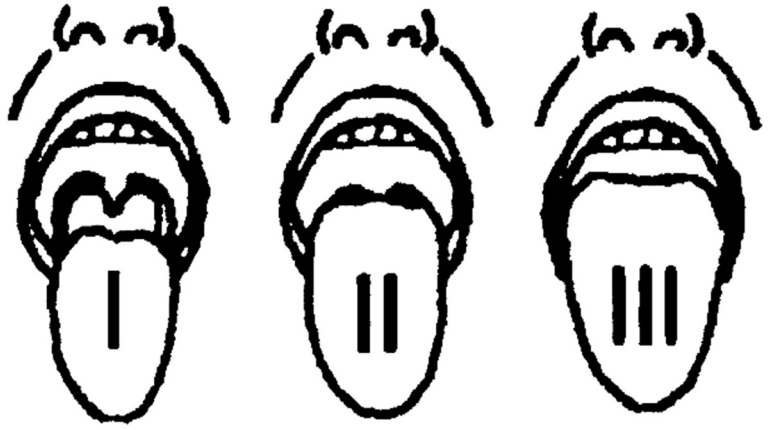


Figure 2 Original Mallampati grades.

1. Changes in the neck due to radiation
2. Presence of beard
3. Snoring
4. Sleep apnoea

Definition of difficult intubation and difficult mask ventilation

El-Ganzouri et al. classified laryngoscopy view after Cormack and Lehane's Class I to IV grading system [21] and used it as a surrogate measure for difficult intubation. In the DIFFICAIR trial, an intubation score is programmed in the DAD based on numbers of attempts and use of equipment (Figure 5). Thus describing the actual circumstances regarding the intubation. An equivalent score is programmed for mask ventilation.

Data registration on the anaesthesia record

The variables in the SARI model are recorded on an appendix to the anaesthesia record either on a pre-printed label adhered to the record form or on a pre-printed supplementary form that is stapled on to the anaesthesia record.

Data registration in the DAD

For all patients the following variables are recorded: i) Preoperative airway assessment (Figure 6); ii) Scheduled airway management (Figure 7); iii) Actual airway management (Figure 5).

The registration of the preoperative airway assessment differs according to group.

SARI group:

A. Predictors included in the SARI model
B. The SARI score
C. Predictors of difficult mask ventilation
D. The anaesthetist's assessment:

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

Is mask ventilation anticipated to be difficult? Yes/No.

CA group:

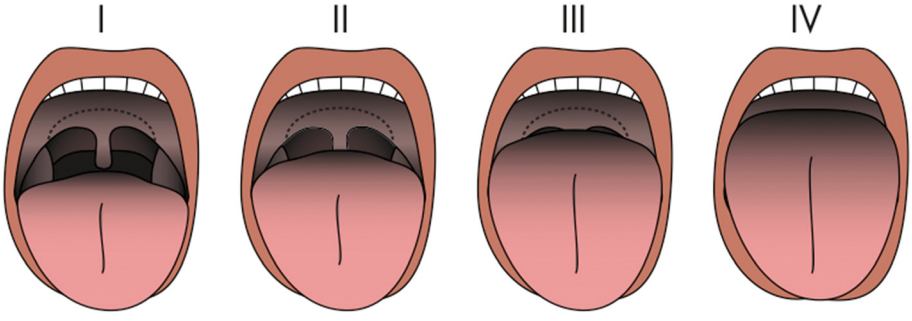


Figure 3 Modified Mallampati classification.

Mallampati I-III		Modified Mallampati I-IV	
Grade	Visualisation	Class	Visualisation
I	Soft palate, uvula, faucial pillars	I	Soft palate, fauces, uvula, faucial pillars
		II	Soft palate, fauces, uvula
II	Soft palate and faucial pillars could be visualizes, but uvula is masked by the base of the tongue	III	Soft palate, base of uvula
III	Only hard palate or part of the soft palate could be visualized	IV	Soft palate not visible at all

Figure 4 Mallampati comparison. The original Mallampati uses three grades of visualisation and the modified Mallampati uses four classes.

The CA group uses variables that are already registered in the DAD.

A. The anaesthetist's assessment:

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

Is mask ventilation anticipated to be difficult? Yes/No.

The seven predictors of difficult intubation contained in the SARI model are registered in the database. Based on these data, the DAD auto-generates a SARI score. In addition, the values of the four predictors of DMV are registered. Despite knowing the SARI score at the pre-operative assessment, the anaesthetist's assessment of anticipated difficulties can differ from the SARI score. Hence, the score is only meant to be indicative of intubation difficulties or not. Following airway management, the actual airway management conditions are finally recorded (Figure 5).

Estimation of sample size

The required number of patients for the detection or rejection of a given effect of the intervention in a cluster randomized trial is calculated by adjusting the required number of patients in a corresponding individually randomized trial with the degree of variation between the clusters (between-cluster variance) [22]. This method is analogue to adjustment with the intra-cluster correlation coefficient [22]. Deviations from the individual sample size estimation are necessary [18] and the calculation can be based on comparison between the groups at cluster level, if the following four conditions are met: i) the intervention is used strictly on cluster level; ii) patients and anaesthetists (intubators) do not migrate between clusters; iii) patients/anaesthetists (intubators) cannot be selected for, or by themselves select/deselect, the intervention; iv) all patients in each cluster are exposed to the intervention and no patient chooses a cluster based on preference for one type of airway assessment.

Actual airway management conditions

Intubation
Regardless of randomization intubation is graded according to the following score. One of the below options is chosen in succession of the intubation procedure, or alternatively failure:

0. Not attempted
1. A maximum of 2 intubation attempts – Only by direct laryngoscopy
2. A maximum of 2 intubation attempts in which other intubation equipment or assistive devices for direct laryngoscopy is used
3. Three intubation attempts or more - Regardless of intubation method
4. Intubation failed despite attempts

Tracheal intubation by direct laryngoscopy is defined as unproblematic by a score = 1 and difficult by a score ≥ 2

Mask ventilation
Regardless of randomization mask ventilation is graded according to the following score. One of the below options is chosen:

0. Not attempted
1. Can be mask ventilated
2. Can be mask ventilated, but with difficulties
3. Impossible to mask ventilate

Difficult mask ventilation is defined as: Inadequate, unstable or requiring two providers, with or without muscle relaxant.
Impossible mask ventilation is defined as: Unable to mask ventilate with or without muscle relaxant.

Figure 5 Actual airway management conditions. Intubation and mask ventilation score in the Danish Anaesthesia Database.

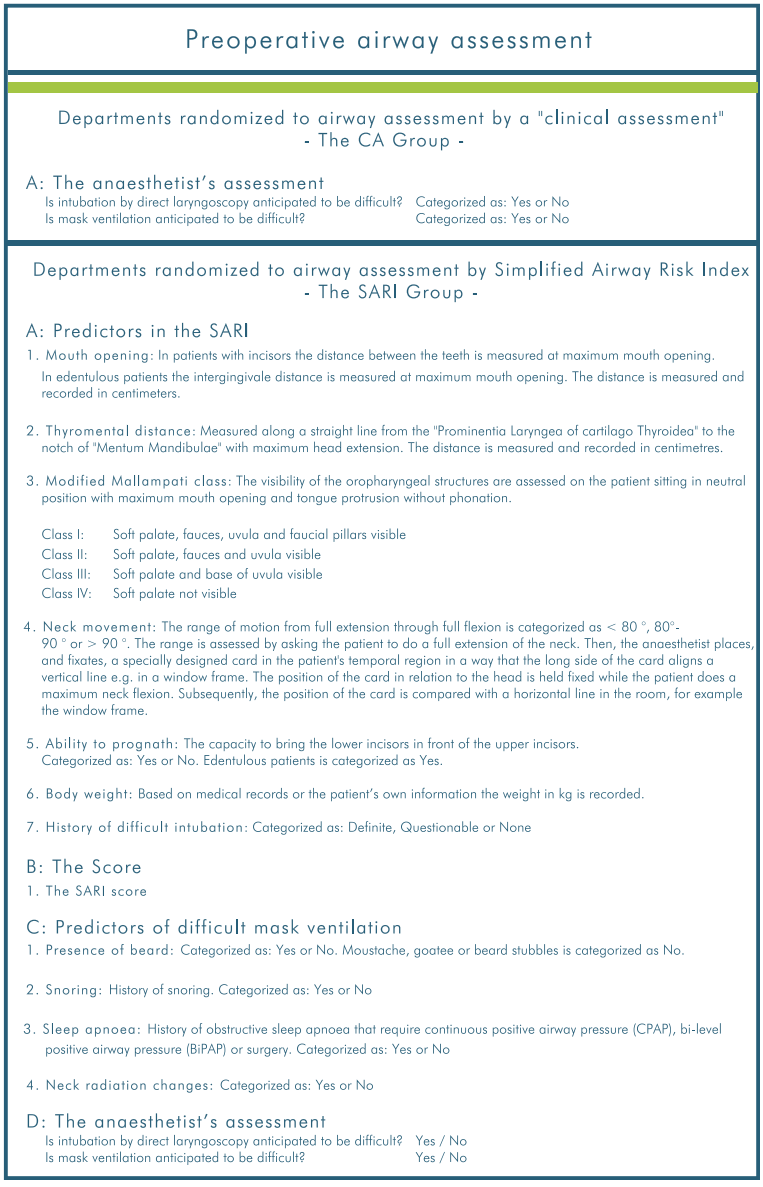


Figure 6 Preoperative airway assessment. Registration of the preoperative airway assessment in the Danish Anaesthesia Database is dependent on the randomization and group allocation.

The estimated number of departments required for inclusion in the trial is based on data extraction from the DAD on patients who had unanticipated difficult tracheal intubations.

There were no previous records of the trial's primary outcome measure, "unanticipated difficult intubation". A baseline study was conducted using data from the DAD generated between 1 January and 1 June 2011. A total of 29 departments met the requirements of cluster size and registration of unanticipated difficult intubations. There were a total of 31,268 intubations or intubation attempts, of which 584 were unanticipated difficult, corresponding to a proportion of 1.87%. We calculated the cluster size and proportions of unanticipated difficult intubation for each department and used this to calculate the "between-cluster variance".

The sample size estimation was further adjusted for the stratification of departments according to their proportions of unanticipated difficult intubation. The estimation was also adjusted according to sample size adjustments in matched cluster trials. We assume that the coefficient of variation, k, is similar in both the CA and SARI groups. Thus, the sample size estimation based on the baseline study data led to k = 0.25, corresponding to an intracluster correlation coefficient of 0.002, and an adjusted average cluster size of 1,611 patients.

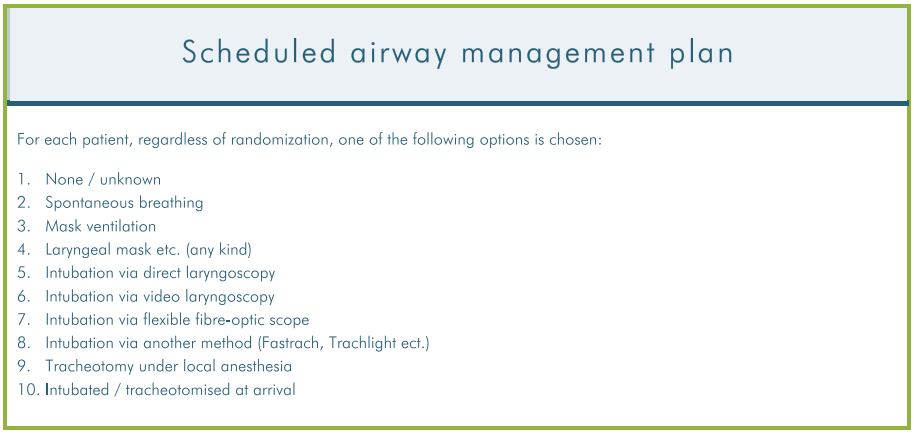


Figure 7 Scheduled airway management. The scheduled airway management plan entered into the Danish Anaesthesia Database.

In the stratified randomization, choosing a power of 80% (power = 1-β, with β being the maximal risk of type 2 error) and a maximal risk of type 1 error of 5% (α = 0.05, two-sided), we intend to be able to detect or reject a 30% relative risk reduction from 1.87% to 1.31%. Given these assumptions, approximately 30 departments are required for the trial: 15 in the SARI group and 15 in the CA group. In this case, it is possible to show that number needed to treat (NNT) is 180 or less. Therefore, we will be able to avoid one unanticipated difficult intubation for every 180 airways assessed by the SARI score instead of a non-specified preoperative airway assessment, if the trial detects a statistically significant difference.

The trial period will be 15 months. We have included 28 departments, randomized and stratified by a proportion of unanticipated difficult intubation less or greater than 2%. The departments are expected to have an average cluster size of approximately 2,500 patients, equalling allocation of approximately 35,000 patients for each trial group.

Statistical analysis

The observed risk factors provide the basis for calculating the SARI score and for preoperative anticipation of a difficult intubation or not. Comparisons between the outcomes of the trial groups will be done on an individual level according to our sample size estimation. In the primary adjusted analysis, the number of patients having an unanticipated difficult (easy) intubation will be compared between the two trial groups with a logistic regression analysis adjusted for stratification variable of baseline proportions of unanticipated difficult (easy) intubation and clustering [23]. The OR for unanticipated difficult (easy) intubation comparing the SARI group with the CA group, and its 95% confidence limits, will be estimated [23]. In an unadjusted analysis, the number of patients having an unanticipated difficult (easy) intubation will be compared between the two trial groups with a χ^2 test. The difference in proportions of unanticipated difficult (easy) intubation will be given with 95% confidence limits. Finally, an adjusted analysis using both stratification variables, the clustering, elective/acute, sex, age, use of neuromuscular blocking agents, and BMI will be performed [7,24].

The accuracy of the SARI score will be compared with the accuracy of the clinical assessment in the CA group on predicting difficult intubation. Additionally, the clinical assessment of the CA group will be compared with the clinical assessment of the SARI group based on the SARI score. That is, anticipations of intubation difficulties based on a clinical assessment only versus anticipations of intubation difficulties based on a clinical assessment while knowing the SARI score. In all analyses, a P value less than 0.05 will be considered statistically significant.

Implementing and sustaining the experimental intervention

Before initiation of the trial, anaesthetists, in departments randomized for the use of the SARI model, were systematically trained in the performance of proper airway assessment according to the SARI score. This ensures uniform and high quality airway assessments [25]. A tutorial film, describing the trial and the preoperative SARI airway assessment in detail, has been produced and was shown to all anaesthetists. In each department, a principal investigator was appointed to ensure individual training of anaesthesiologists in correct airway assessment at trial start and again after 6 months. All new employees also receive this training. A short description of the DIFFICAIR trial is included in the introduction material for new physicians and nurses on intervention departments. Posters were placed and flyers made available describing the SARI model. A card that fits uniform

pockets was produced. The card includes a ruler and a protractor to facilite airway assessment.

A website, www.difficair.com, containing all information including the tutorial film, PowerPoint presentations and other tools for education, was programmed. Different material is available on the website for the SARI- and the CA group. Different access is granted via different passwords.

Implementing and sustaining the control intervention

For the departments randomized to the CA group, there will be no changes in registration of data in the DAD compared to usual standards. On the anaesthesia record (or directly into the DAD), the preoperative airway assessment and the scheduled airway management are recorded before the anaesthesia and actual airway management begins.

For both groups, a mask ventilation and/or an intubation score (Figure 5) is registered on the anaesthesia record or directly in the DAD during anaesthesia and immediately after airway management.

Regardless of trial group, the DAD registration is performed during or immediately after the end of anaesthesia. Data is entered via a computer workstation with an Internet connection to DAD. The anaesthetist who performed the airway management or the anaesthetist who completed the anaesthesia performs the registration.

In case of technical problems with the DAD, e.g., loss of Internet connection, a paper form corresponding to the electronic interface in DAD is used. Relevant personnel, e.g., a secretary, subsequently enter data from the form into the DAD at restoration of Internet connection.

After trial completion, data will be retrieved from the DAD, guarantying patients' anonymity, following the rules of the Danish data protection agency. Anonymous data will be made accessible by other researchers through the Danish Data Archive.

Data monitoring

Through the trial period, the degree of data completeness will be continuously monitored for each department. In case of a declining percentage of registration the principal investigator in the corresponding department will be contacted in order to restore the registration rate. The investigator group is blinded for all outcome measures during the trial period.

Handling of incomplete data

Missing data exceeding a rate of 5% and with a statistical significant Little's test, precluding analyses on the data set of complete cases, will be handled statistically through multiple imputation [26-28].

Trial registration and ethics

The trial is a database research project involving registration of variables that are already being observed in the involved departments to varying degrees. The trial is without risks, side effects or inconvenience for the patient, and the trial protocol includes no specific dictation on airway management.

The Scientific Ethics Committee of Copenhagen County consents that the protocol should not be reported to the committee system (Journal No.: H-3-2012-FSP2). Individual informed consent from the patients is not necessary, which is essential for the feasibility. However, informed consent from every participating department by the Head of Department was acquired. The trial is approved by The Danish Data Protection Agency (j.nr.: 2007-58-0015/HIH-2011-10, I-Suite nr: 02079) and is registered at http://www.clinicaltrials.gov (NCT01718561).

Publications

The protocol is written according to the SPIRIT 2013 recommendations [29]. Results of the trial will be reported according to the CONSORT statement for cluster randomized trials [18] and the STROBE criteria [30].

Manuscripts are written for publication in international peer reviewed journals. First author is Anders K. Nørskov, MD, Department of Anaesthesiology, Nordsjællands Hospital – Hillerød. Additional authors are Jørn Wetterslev, Chief Physician, MD, PhD, Copenhagen Trial Unit, Rigshospitalet; Charlotte V. Rosenstock, Consultant, MD, PhD and Lars H. Lundstrøm, MD, PhD, both from the Department of Anaesthesiology, Nordsjællands Hospital – Hillerød. The DIFFICAIR steering committee will grant additional authorship in accordance with the Vancouver rules and all trial site investigators are acknowledged with co-authorships.

Based on the trial results and international literature we hope to contribute to a national recommendation for preoperative airway assessment and its subsequent implementation.

The method used in the study is "state of the art" for testing an implementation of a recommendation [31,32]. All manuscripts will be submitted for publication to international peer-reviewed journals, published in annual reports from the DAD, and presented at national and international congresses.

Side studies will be allowed in accordance with the steering committee.

Timeline

2011: Applications for funding. Acceptance from ethical committee. Baseline study and sample size calculation.

2012: Applications for funding. Written consent to randomization from 28 departments of anaesthesia.

Revision and programming of the DAD. Education of intervention departments.

End 2012: First patient inclusion.

End 2013: Last patient inclusion.

Early 2014: Data analysis. Writing and submission of main manuscripts for publication.

Collaborations and finances

The trial is done in collaboration between the Danish Anaesthesia Database; Department of Anaesthesiology, Copenhagen University Hospital, Nordsjællands Hospital – Hillerød; Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen University Hospital, Rigshospitalet; and 28 Danish departments of anaesthesia. All participating departments provided written consent for inclusion and randomization for the trial. The DAD steering committee supports the DIFFICAIR trial and the data extraction is done in agreement with the committee.

The investigators have no financial ties to private companies or foundations and no potentially conflicting interests in the project.

The study is fully funded by the Tryg Foundation; the Research foundation at Copenhagen University Hospital, Nordsjællands Hospital – Hillerød; DASAIMs fund; and resources at local trial sites. None of the funding sources has any influence on protocols, data handling or publications.

Discussion and perspective

It is innovative to use a national clinical database as the basis for a randomized clinical trial. The method can serve as a precedent for implementation of evidence-based recommendations and database registrations.

The trial will forward understanding of how to predict and reduce the unanticipated difficult intubation and mask ventilation and how to produce evidence-based recommendations for airway assessment nationally and internationally.

Trial status

The trial was initiated on October 1, 2012 through DAD recording in all intervention and control departments. Two control departments still await connection to the DAD registry. Nevertheless, they are expected to meet the minimum inclusion criteria before trial termination.

Patient recruitment was ongoing at time of submission of the manuscript.

The trial ends at the end of 2013.

Abbreviations

ASA: American Society of Anesthesiologists; BMI: Body mass index; CA: Clinical assessment; DAD: Danish Anaesthesia Database; DMV: Difficult mask ventilation; NNT: Number needed to treat; SARI: Simplified Airway Risk Index.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AN participated in the study design, contributed to the use of study methodology, carried out the statistical calculations and drafted the manuscript. CR participated in the design of the study and helped to draft the manuscript. JW participated in the design of the study, contributed to the use of study methodology and statistical calculations, and helped to draft the manuscript. LL participated in the design of the study and the use of study methodology, and helped with the statistical calculations and the drafting of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The TRYG foundation; The Danish Anaesthesia Database; The Nordsjællands Hospital-Hillerød research council; The Capital Regions funds for research; DASAIMs fund. Thanks to Jakob L. Thomsen, MD, for assistance on development of educational materials. Thanks to Grethe Astrup, MD, Consultant (Århus University Hospital) and to Arash Afshari, MD, PhD, Consultant (Juliane Marie Center, Copenhagen University Hospital, Rigshospitalet) for additional advice on the protocol. This protocol is published on behalf of the DIFFICAIR trial group.

Funding

The trial is funded by independent grants.

Received: 21 May 2013 Accepted: 4 October 2013
Published: 23 October 2013

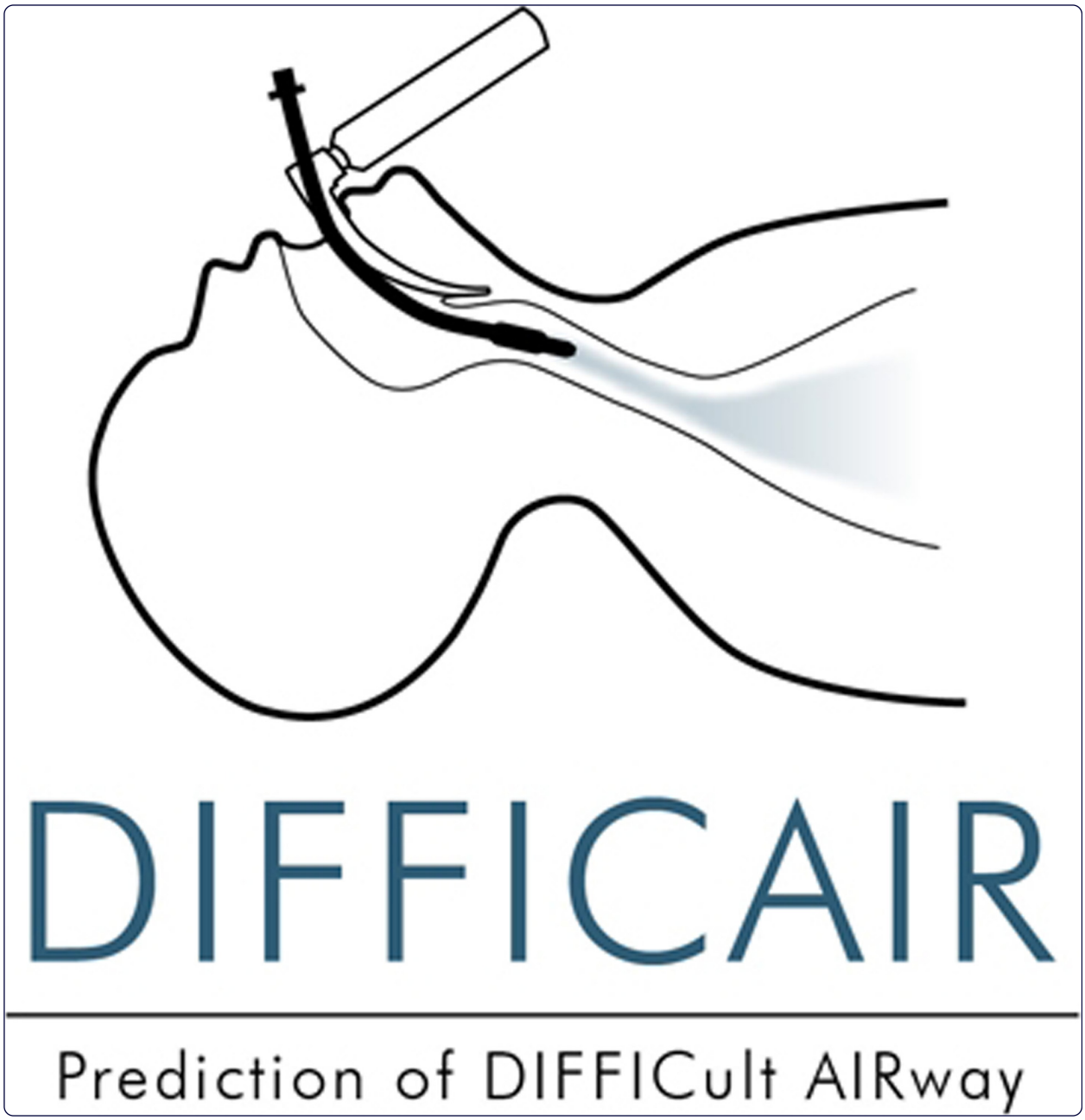
References

1. Cooper GM, McClure JH: **Anaesthesia chapter from saving mothers' lives; reviewing maternal deaths to make pregnancy safer.** *Br J Anaesth* 2008, **100:**17–22.
2. Hove LD, Steinmetz J, Christoffersen JK, Møller A, Nielsen J, Schmidt H: **Closed claims analysis of deaths related to anesthesia in the period 1996–2004 registered by the Danish patient insurance association.** *Anesthesiology* 2007, **106:**2780–2781.
3. 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.
4. Rosenstock C, Moller 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:**53–58.
5. 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:**659–667.
6. 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.
7. 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.
8. Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Gätke MR, Wetterslev J: **Avoidance of neuromuscular blocking agents may increase the risk of difficult tracheal intubation: a cohort study of 103,812 consecutive adult patients recorded in the Danish anaesthesia database.** *Br J Anaesth* 2009, **103:**283–290.
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:**1081–1088.
10. Shiga T, Wajima Z, Inoue T, Sakamoto S, Sakamoto A: **Predicting difficult intubation in apparently normal patients.** *Anesthesiology* 2005, **103:**429–437.
11. 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.
12. 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.

13. 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:**251–270.
14. 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.
15. Henderson JJ, Popat MT, Latto IP, Pearce AC: **Difficult airway society guidelines for management of the unanticipated difficult intubation.** *Anaesthesia* 2004, **59:**675–694.
16. 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:**266–271.
17. 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.
18. Campbell MK, Elbourne DR, Altman DG: **CONSORT statement: extension to cluster randomised trials.** *BMJ Br Med J* 2004, **328:**702–708.
19. Hahn S, Puffer S, Torgerson DJ, Watson J: **Methodological bias in cluster randomised trials.** *BMC Med Res Methodol* 2005, **5:**10.
20. Samsoon GLT, Young JRB: **Difficult tracheal intubation: a retrospective study.** *Anaesthesia* 1987, **42:**487–490.
21. Cormack RS, Lehane J: **Difficult tracheal intubation in obstetrics.** *Anaesthesia* 1984, **39:**1105–1111.
22. Hayes RJ, Bennett S: **Simple sample size calculation for cluster-randomized trials.** *Int J Epidemiol* 1999, **28:**319–326.
23. Kahan BC, Morris TPT: **Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis.** *BMJ Br Med J* 2012, **345:**e5840.
24. Lundstrøm LH: **Detection of risk factors for difficult tracheal intubation.** *Dan Med J* 2012, **59:**B4431.
25. Rosenstock C, Ostergaard D, Kristensen MS, Lippert A, Ruhnau B, Rasmussen LS: **Residents lack knowledge and practical skills in handling the difficult airway.** *Acta Anaesthesiol Scand* 2004, **48:**1014–1018.
26. Rubin DB: *Multiple Imputation for Nonresponse in Surveys.* Wiley online library; 2008 [http://onlinelibrary.wiley.com/book/10.1002/9780470316696].
27. Schafer JL: **Multiple imputation: a primer.** *Stat Methods Med Res* 1999, **8:**3–15.
28. Schafer JL, Graham JW: **Missing data: our view of the state of the art.** *Psychol Methods* 2002, **7:**147–177.
29. Chan AW, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, Dickersin K, Hróbjartsson A, Schulz KF, Parulekar WR, Krleza-Jeric K, Laupacis A, Moher D: **SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials.** *BMJ* 2013, **346:**e7586.
30. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP: **The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies.** *J Clin Epidemiol* 2008, **61:**344–349.
31. 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.
32. Grimshaw JM, Eccles MP: **Is evidence-based implementation of evidence-based care possible?** *Med J Aust* 2004, **180**(6 Suppl):S50–S51.

doi:10.1186/1745-6215-14-347
Cite this article as: Nørskov *et al.*: 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.

TRIALS



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

Nørskov *et al.*

BioMed Central

Nørskov *et al. Trials* 2014, **15**:173
http://www.trialsjournal.com/content/15/1/173

TRIALS

UPDATE Open Access

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

Anders Kehlet Nørskov[1,2*], Lars Hyldborg Lundstrøm[1], Charlotte Vallentin Rosenstock[1] and Jørn Wetterslev[2]

Abstract

Background: Preoperative airway assessment in Denmark is based on a non-specific clinical assessment left to the discretion of the responsible anesthesiologist. The DIFFICAIR trial compares the effect of using a systematic and consistent airway assessment versus a non-specific clinical assessment on the frequency of unanticipated difficult airway management.

To prevent outcome bias and selective reporting, we hereby present a detailed statistical analysis plan as an amendment (update) to the previously published protocol for the DIFFICAIR trial.

Method/Design: The DIFFICAIR trial is a stratified, parallel group, cluster (cluster = department) randomized multicenter trial involving 28 departments of anesthesia in Denmark randomized to airway assessment either by the Simplified Airway Risk Index (SARI) or by a usual non-specific assessment. Data from patients' preoperative airway assessment are registered in the Danish Anesthesia Database. An objective score for intubation grading the severity, that is the severity of the intubations, as well as the frequency of unanticipated difficult intubation, is measured for each group.

Primary outcome measures are the fraction of unanticipated difficult and easy intubations.

The database is programmed so that the registration of the SARI is mandatory for the intervention group but invisible to controls.

Data recruitment was commenced in October 2012 and ended *in ultimo* December 2013.

Conclusion: We intend to increase the transparency of the data analyses regarding the DIFFICAIR trial by an *a priori* publication of a statistical analysis plan.

Trial registration: ClinicalTrials.gov: NCT01718561.

Keywords: Statistical analysis plan, Cluster randomized trial, Airway management, Cluster analysis, Difficult intubation

Introduction

The difficult airway management trial (DIFFICAIR) is a stratified, parallel group, cluster (cluster = department) randomized and multicenter trial involving 28 departments of anesthesia in Denmark. The DIFFICAIR trial compares the effect of two regimens of preoperative airway assessment on the frequency of unanticipated difficult airway management.

Prediction of difficult airway management remains a pivotal challenge in anesthesia. Difficult tracheal intubation and difficult mask ventilation may cause serious patient complications [1-6]. By allocating experienced personnel and relevant equipment, better prediction of difficult airway management may reduce complications and, thereby, associated morbidity and mortality. There is no single predictor that is sufficiently valid in predicting difficult tracheal intubation [7-12]. However, several studies show that by combining multiple predictors of difficult tracheal intubation, the positive and the negative predictive value of the assessment increases [12]. In Denmark as well as internationally, there is no clear recommendation on how to perform airway assessment. Consequently, airway assessment in Denmark is based exclusively on the individual anesthesiologist's preoperative clinical assessment. However, it is poorly documented how accurately this clinical assessment predicts actual airway management conditions.

The 'Simplified Airway Risk Index' (SARI) [13] is based on a multivariable model for airway assessment described

* Correspondence: anderskehlet@hotmail.com
[1]Department of Anaesthesiology, Nordsjællands Hospital, Copenhagen University Hospital, Hillerød, Capital region of Denmark 3400, Denmark
[2]Copenhagen Trial Unit, Center for Clinical Intervention Research, Copenhagen University Hospital, Rigshospitalet, Copenhagen 2100, Denmark

by El-Ganzouri and colleagues enabling an estimation of the likelihood of a difficult direct laryngoscopy. The SARI contains seven individual predictors for a difficult direct laryngoscopy, each given a weighted score of 0 to 1 or 0 to 2. A summarized value of the SARI score > 3 indicates that a future direct laryngoscopy will be difficult. It is unknown, whether the SARI score predicts difficult intubation better or worse than a clinical assessment. The rationale for this trial was to prospectively compare the effect of the SARI with an non-specified clinical airway assessment on the frequency of unanticipated difficult airway management.

The target population was adult patients undergoing anesthesia. Twenty-eight departments of anesthesia were randomized to one of two groups. Intervention departments used the SARI score for preoperative airway assessment. The intervention group additionally did an assessment of risk factors for difficult mask ventilation as described by Kheterpal and colleagues [14-16]. Departments in the control group continued normal practice of preoperative airway assessment. All data were registered in the Danish Anesthesia Database (DAD). A more detailed trial protocol describing background, design and rationale has been published in *Trials* [17].

In order to prevent outcome reporting bias and results based on data-driven analysis, it is encouraged to prospectively publish a trial protocol [18,19]. The same argument applies for a prospective publication of a statistical analysis plan. Concordantly, the International Conference on Harmonization (ICH) of Good Clinical Practice (GCP) recommends that clinical trials are analyzed according to a pre-specified plan [19].

Objective

The primary aim of the DIFFICAIR trial is to compare the effect of using a systematic airway assessment with a standard clinical airway assessment on the frequency of unanticipated difficult airway management. The null hypothesis is:

- There is no difference in the proportion of unanticipated difficult intubations when the preoperative airway assessment is based on the SARI score compared with a preoperative airway assessment based on the individual anesthesiologist's assessment.

The alternative hypothesis is:

- The use of a systematic SARI airway assessment, registration of the SARI and risk factors for difficult mask ventilation, and continuous education in airway assessment will reduce the relative risk of a difficult intubation with 30%, corresponding to a number needed to treat (NNT) of 180 patients.

Methods

This analysis plan has been written while the data collection from the DIFFICAIR trial was on-going and trial data non-accessible. The data analysis of the main publication will follow this plan. The statistical analysis was approved by the DIFFICAIR steering committee on 29 December 2013. The last day of data collection was 31 December 2013. The involved departments were given one additional month to ensure registration of all patients in the Danish Anesthesia Database. On 31 January 2014, the database was locked and data extracted. The statistical analysis plan was published on (www.clinicaltrials.gov) before the last data entry and before data was extracted and data management commenced.

The DIFFICAIR trial protocol has been written according to the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) guidelines and has been public on (www.difficair.com) since the beginning of the trial and is registered at (www.clinicaltrials.gov) (NCT01718561). The Danish Anesthesia Database and the Danish Society of Anaesthesiology and Intensive care Medicine (DASAIM) endorsed the trial.

The trial is carried out in accordance with the Helsinki declaration. The Scientific Ethics Committee of Copenhagen County has declared that it is regarded as a quality assurance project and thus should not be reported to the committee system (Journal number: H-3-2012-FSP2). Further, the need for individual patient consent was waived. The trial is approved by The Danish Data Protection Agency (Journal number: 2007-58-0015/HIH-2011-10, I-Suite number: 02079). The reporting of the trial will be in accordance with the CONSORT 2010 statement: extension to cluster randomised trials [20].

Randomization and sample size

Our sample size calculation was based on an adjustment for the stratification and the cluster randomized design [21,22]. Since there are no previous records of the trial's primary outcome measure, 'unanticipated difficult intubation' a baseline study was conducted based on data from the DAD. In order to reject or detect a 30% relative risk reduction in the proportions of unanticipated difficult intubation between the intervention group and the control group approximately 30 departments were required in a 15 months period. Calculations were based on a maximum risk of type 1 error of 5% and risk of type 2 error of maximum 20% (80% power).

A total of 28 departments were included and randomized 1:1 using a computer generated list. The sample size calculation was based on an average cluster size of 1,611 patients. We estimated the average cluster size in the DIFFICAIR trial to approximately 2,500 patients, giving a total of 70,000 included patients during the trial period. The enhanced sample size allows for a potentially

slight loss of clusters according to the power calculation, from 30 to potentially 26. Our sample size estimation may be of a conservative nature, calling for more clusters than necessary [23].

Populations

The DIFFICAIR trial focuses on two essential elements of airway management which are tracheal intubation by direct laryngoscopy and mask ventilation. This statistical analysis plan will address analysis of the data regarding tracheal intubation. Data analysis regarding prediction of difficult mask ventilation will be handled in an analogous way, but will not be further elaborated in the present paper.

The part of the DIFFICAIR trial regarding prediction of difficult intubation comprises two populations; 1) patients that were primarily attempted intubated by direct laryngoscopy; 2) patients that were primarily attempted intubated by direct laryngoscopy (population 1) plus patients anticipated to be difficult to intubate and therefore scheduled for and intubated with an advanced method (for example, video laryngoscopic or fiber optic intubation).

The results of population 1 and 2 will be presented in one publication. Due to the extent of data, further publications presenting data from the DIFFICAIR trial will follow, but further elaboration on data analysis exceeds the content frame of this paper.

Adjusting and stratification variables

Each cluster (department) was randomized to a control or intervention group, making this the intervention group indicator. The trial site may account for further intervention heterogeneity and will be used for adjustment in the analysis of the intervention effect. Further, a stratification variable that grouped the departments according to whether the proportion of unanticipated difficult intubation at baseline was ≥ or < 2% will be used for adjustment according to recent evidence of increased power in the analysis of stratified trials [22].

Assumed confounding covariates

We define age; gender; ASA classification; emergency/elective procedure; Body Mass Index (BMI); and use of neuromuscular blocking agents as covariates that are possible confounders, necessitating adjusted analyses of the primary outcome and pre-defined subgroup analyses.

Primary outcomes

The primary outcome measures are:

1. The fraction of unanticipated difficult intubations = all intubations with unanticipated difficulties (False negative)/all patients primarily (attempted) intubated by direct laryngoscopy.
2. The fraction of unanticipated easy intubations = all intubations with anticipated difficulties that were easy (False Positive)/all patients primarily (attempted) intubated by direct laryngoscopy.

The two primary outcomes are linked and simultaneous low fractions are desirable for the optimal prediction of a difficult intubation.

Secondary outcomes

1. 48-hour mortality.
2. 30-day mortality.
3. The fraction of anticipated difficult intubations planned for, and intubated by an advanced method/all patients (attempted) intubated.
4. The fraction of unanticipated difficult intubations (False Negative)/all difficult intubations ((False negative) + (True Positive)).
5. Sensitivity of the prediction of a difficult/easy intubation.
6. Specificity of the prediction a difficult/easy intubation.
7. Predictive value of a positive prediction of difficult/easy intubation.
8. Predictive value of a negative prediction of difficult/easy intubation.
9. Positive Likelihood Ratio = (Sensitivity/(1-Specificity)).
10. Negative Likelihood Ratio = ((1-Sensitivity)/Specificity).
11. The Receiver Operating Characteristic (ROC) curve. A graphical representation of sensitivity as a function of (1-Specificity).

Outcomes 5 to 10 are measured for both intervention groups.

Outcome 11 will be measured on relevant non-binary predictors.

Datapoints

Baseline covariates

Individual level:

1. Sex
2. Age
3. Height
4. Weight
5. BMI
6. American Society of Anesthesiologists (ASA) Classification
7. Use of neuromuscular blocking agents
8. Hospital unit
9. Region
10. Anticipated difficult tracheal intubation
11. Anticipated difficult mask ventilation
12. Scheduled airway
13. Priority: emergency/elective

14. Surgical procedure codes
15. Intubation score
16. Mask ventilation score.

Intervention covariates

1. Mouth opening
2. Thyro-mental distance
3. Modified Mallampati classification
4. Jaw protrusion
5. Neck mobility
6. Previous difficult airway management
7. Number of completed risk factors
8. The calculated SARI score
9. Dichotomized SARI score (<or ≥ 4)
10. Snoring
11. Sleep apnoea
12. Presence of beard
13. Changes in the neck due to radiation.

Cluster level summaries

1. Mean cluster size
2. Mean number of intubated patients
3. Fraction of private hospitals
4. Mean fraction of unanticipated difficult intubation
5. Mean fraction of unanticipated easy intubation
6. Age
7. BMI
8. ASA classification.

Definition of difficult intubation

In the DAD, an intubation score is programmed based on numbers of intubation attempts and use of equipment.

1. A maximum of two intubation attempts - only by direct laryngoscopy.
2. A maximum of two intubation attempts in which other intubation equipment or assistive devices for direct laryngoscopy is used (for example, video laryngoscope).
3. Three intubation attempts or more - regardless of intubation method.
4. Intubation failed despite attempting.

Tracheal intubation by direct laryngoscopy is pre-defined in the DAD as easy by a score = 1 and difficult by a score ≥ 2. In our primary analyses and sample size calculation we employ the same definition.

General analysis principles

1. Unless otherwise stated, all main analyses will compare the two intervention groups using intention-to-treat (ITT) [24].
2. 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 [25-27]. Analyses will be based on individual patient level data but clustering of patients and the stratification variable will be accounted for in a generalized estimating equation.
3. In all analyses, a maximum level of 5% (two-sided) type 1 error will be regarded as statistically significant unless otherwise stated.
4. Main analyses will be according to ITT adjusted for cluster and stratification variables. Sensitivity analyses will be performed adjusted and unadjusted for the prior listed potential confounding covariates. We will discuss if results differ from the main analyses. The conclusion of the trial will be based on the primary analyses.
5. Test of interaction will be applied for subgroup analyses.
6. Risks are reported as relative risks and odds ratios. When relative risks are calculated from odds ratios with 95% confidence interval (CI) it will be done according to Zhang and Yu [28].
7. For missing data exceeding a rate of 5%, and with a statistical significant Little's test, indicating that the missing data is not a completely random sample of the total data, point estimates with 95% CI will be calculated using a worst/best case scenario imputation on the missing values. If the imputation of a worst/best case scenario implies different conclusions, multiple imputations will be performed on the missing values assuming missingness at random [29]. Unadjusted and complete case analyses will also be presented.
8. In order to avoid rejecting a true null hypothesis we will address the problem of multiplicity by Bonferroni adjustments on the secondary outcome measures. If unadjusted analyses are insignificant ($P > 0.05$), Bonferroni adjustments will not be applied. In case the adjustment changes an unadjusted significant *P*-value to a non-significant *P*-value, this will be discussed.
9. To ensure complete objectivity, the author (AN) will be blinded for the intervention group in the primary outcome analysis and, as far as this is possible, for analyses of secondary outcomes. However, analyses of the predictive properties of the SARI will require un-blinding of AN. After data collection, a third party data manager will generate a complete dataset with blinded coding of the intervention groups and other variables possibly revealing the intervention. The statistician performs the primary outcome analysis on this data set. If the primary outcome differs between groups, we will construct different

conclusions reflecting the results, considering that significant differences of the intervention could both be of benefit or harm. After writing the conclusions, we will uncover the code of the blinding, and subsequently the correct conclusion will be employed [30].

Statistical analyses

Trial profile

The flow of study participants will be displayed in a Consolidated Standards of Reporting Trials (CONSORT) diagram at a cluster level and at individual level. The number of clusters fulfilling the inclusion criteria, and the number of clusters included in primary and secondary analyses, will be presented. The number of patients who fulfilled study inclusion criteria as well as the number included in the primary and secondary analyses will be reported. Reasons for exclusions of clusters and patients in the primary and secondary analyses will be reported.

Primary outcome

Frequencies and percentages per group will be reported with a 95% CI. The primary outcome is presented as odds ratios and relative risk ratios.

The primary analysis of the primary outcome will be adjusted for the stratification- and the cluster-variable performed according to the ITT principle including patients that met the inclusion- and not the exclusion-criteria. A generalized estimating equation will be used. Intervention group and stratification variable are regarded as fixed effects and trial site is regarded as random effects in the model. We will test the robustness of the results by repeating the analyses with a mixed effects model and finally with a standard *t*-test comparing the means of the outcome at department level in each intervention group.

The first sensitivity analysis of the primary outcome will be adjusted for the stratification- and cluster-variables as well as baseline covariates assumed as confounders incorporated in a generalized estimating equation.

In the second sensitivity analysis of the primary outcome, we will employ a different cut-off value for difficult intubation using ≥ 3 instead of ≥ 2 as the definition of difficult intubation.

Further sensitivity analyses of the primary outcome will compare the patients in the control group that met the inclusion- and not the exclusion-criteria with patients in the intervention group who received the protocoled intervention. That is, a *per protocol* analysis of control group versus the subgroup in the intervention group that had a sufficiently registered SARI. Interaction test will be performed in the intervention group between patients receiving sufficient/insufficient SARI registration.

Secondary outcomes

Frequencies, proportions, percentages, odds and risk ratios are presented with a 95% CI for each group. A chi-squared test is used to assess the effect of the intervention on binary outcomes. For categorical outcomes and the adjusted analyses, logistic regression analysis or generalized estimating equations will be performed.

Baseline comparisons of patient characteristics

Baseline characteristics are presented for each intervention group. Frequencies, proportions and percentages will be used to summarize discrete variables. In case of missing values, percentages are presented with the actual denominator and otherwise calculated according to the number of participating patients. Continuous variables are summarized using standard measures of central tendency and dispersion using either mean ± SD for data with normal distribution or median and interquartile range for non-normally distributed data.

Baseline comparisons of cluster characteristics

Cluster characteristics are presented for each group, control and intervention. Unless otherwise stated, data will be presented as means with SD for data with normal distribution or median and interquartile range for non-normally distributed data.

Outline of figures and tables

The first figure will be a CONSORT flow chart on individual patient level and cluster level. A second figure will illustrate the SARI score and tutorial instruments. A third figure will demonstrate the registration in the DAD, including the intubation score. A fourth figure will present baseline data from each intervention group on individual and cluster level and a fifth figure will be outlining the main outcome results for each intervention group.

Discussion

In order to avoid outcome reporting bias and data-driven results this paper presents the detailed statistical analysis plan for the main publication of the DIFFICAIR trial. The DIFFICAIR trial raises two important questions, which are: is it possible via the intervention to reduce the frequencies of difficult intubation and/or difficult mask ventilation? This plan only addresses the statistical analyses of the population of intubated patients because our sample size calculations were based on this population. Secondly, the SARI was developed as a prediction tool for difficult intubation. Finally, the extent of data necessitates several publications.

By adjusting our primary outcome analysis for different design variables, such as clustering and stratification, we strive to eliminate inflated type 1 error rates as a consequence of the trial design. A generalized estimating

equation is applied based on an evaluation of each variable as having random or fixed effects [31,32].

When multiple comparisons are performed between two groups, you may risk accepting an intervention effect erroneously (type 1 error). There are several approaches that deal with multiple testing. We will employ Bonferroni adjustments on the secondary outcome measures in order to evaluate, identify and discuss dubious significant outcomes that may be due to statistical multiplicity.

The value of a diagnostic test is usually presented as sensitivity and specificity. We have chosen (1 - total accuracy), that is the proportion of unanticipated difficult intubations (False Negative, FN) and the proportion of unanticipated easy intubations (False Positive, FP). Both scenarios are of clinical relevance since the FNs are at risk of hypoxia, increased morbidity and even death, while the FPs are at risk of being imposed unnecessary discomfort by, for example, awake intubation. At the same time, both the FNs and FPs can take up unnecessary resources. Sensitivity and specificity are more difficult to interpret intuitively. Consequently, we chose to present more transparent primary outcomes. Using proportions of unanticipated difficult intubation allowed us to perform a baseline cohort study, on which we based our sample size and power calculations.

By publishing this paper, where we pre-specify our methods and analyses, it is our hope that the results from the DIFFICAIR trial will be as transparent and robust as possible.

Conclusion

This paper presents the principles of analyses of the main outcomes in the DIFFICAIR trial for the first publication based on patients who underwent intubation. Our approach aims to minimize the risk of data-driven results and outcome reporting bias.

Abbreviations

ASA: American Society of Anesthesiologists; BMI: Body Mass Index; CI: confidence interval; DAD: Danish Anesthesia Database; FN: False Negative; FP: False Positive; GCP: Good Clinical Practice; ICH: International Conference on Harmonization; ITT: intention-to-treat; NNT: number needed to treat; ROC: Receiver Operating Characteristic; SARI: Simplified Airway Risk Index; SD: standard deviation; SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials.

Competing interests

The authors declare no financial or non-financial competing interests.

Authors' contribution

AN, LL, and JW proposed the statistical analysis plan. AN drafted the manuscript. AN, LL, CR, and JW participated in the design of the trial. AN conducted the coordination of the trial. AN, LL, CR, and JW read, amended and approved the statistical analysis plan and the final manuscript.

Acknowledgements

The DIFFICAIR trial is funded by independent grants from The TRYG foundation; The Danish Anesthesia Database; Nordsjællands Hospital-Hillerød research council; The Capital Regions funds for research; DASAIMs fund. This analysis plan is published on behalf of the DIFFICAIR trial group.

Received: 3 January 2014 Accepted: 8 May 2014
Published: 15 May 2014

References

1. Cooper GM, McClure JH: **Anaesthesia chapter from Saving mothers' lives; reviewing maternal deaths to make pregnancy safer.** *Br J Anaesth* 2008, **100:**17–22.
2. Hove LD, Steinmetz J, Christoffersen JK, Møller A, Nielsen J, Schmidt H, Moller A: **Analysis of deaths related to anesthesia in the period 19962004 from closed claims registered by the Danish Patient Insurance Association.** *Anesthesiology* 2007, **106:**675–680.
3. 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.
4. 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:**53–58.
5. 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:**127–132.
6. Cook T, Woodall N, Frerk C: **Major complications of airway management in the UK:results of the Fourth National Audit Project of the royal college of anaesthesiologists and the difficult airway society. Part 1: Anaesthesia.** *Br J Anaesth* 2011, **106:**617–631.
7. 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:**659–667.
8. 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.
9. 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.
10. Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Gätke MR, Wetterslev J: **Avoidance of neuromuscular blocking agents may increase the risk of difficult tracheal intubation: a cohort study of 103,812 consecutive adult patients recorded in the Danish Anaesthesia Database.** *Br J Anaesth* 2009, **103:**283–290.
11. 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:**1081–1088.
12. Shiga T, Wajima Z, Inoue T, Sakamoto S, Sakamoto A: **Predicting difficult intubation in apparently normal patients.** *Anesthesiology* 2005, **103:**429–437.
13. 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.
14. 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.
15. 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.
16. Kheterpal S, Healy D, Aziz M: **Incidence, predictors, and outcome of difficult mask ventilation combined with difficult laryngoscopy: a report from the Multicenter Perioperative Outcomes Group.** *Anesthesiology* 2013, **6:**1–10.
17. 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.
18. 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:**200–207.
19. Dwan K, Gamble C, Williamson PR, Altman DG: **Reporting of clinical trials: a review of research funders' guidelines.** *Trials* 2008, **9:**66.
20. Campbell MK, Piaggio G, Elbourne DR, Altman DG: **Consort 2010 statement: extension to cluster randomised trials.** *BMJ* 2012, **345:**e5661.

21. Hayes RJ, Bennett S: **Simple sample size calculation for cluster-randomized trials.** *Int J Epidemiol* 1999, **28:**319–326.
22. Kahan B, Morris T: **Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis.** *BMJ* 2012, **345:**e5840.
23. Crespi CM, Wong WK, Wu S: **A new dependence parameter approach to improve the design of cluster randomized trials with binary outcomes.** *Clin Trials* 2011, **8:**687–698.
24. Fergusson D, Aaron SD, Guyatt G, Hébert P: **Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis.** *BMJ* 2002, **325:**652–654.
25. Klar N, Donner A: **Current and future challenges in the design and analysis of cluster randomization trials.** *Stat Med* 2001, **20:**3729–3740.
26. Hayes RJ, Moulton LH: *Cluster Randomised Trials. Pharmaceutical Statistics, vol 11, Issue 1.*
27. Kahan BC, Morris TP: **Assessing potential sources of clustering in individually randomised trials.** *BMC Med Res Methodol* 2013, **13:**58.
28. Zhang J, Yu KF: **What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes.** *JAMA* 1998, **280:**1690–1691.
29. Schafer JL: **Multiple imputation: a primer.** *Stat Methods Med Res* 1999, **8:**3–15.
30. Gøtzsche PC: **Blinding during data analysis and writing of manuscripts.** *Control Clin Trials* 1996, **17:**285–290.
31. Peters T: **Comparison of methods for analysing cluster randomized trials: an example involving a factorial design.** *Int J Epidemiol* 2003, **32:**840–846.
32. 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.

doi:10.1186/1745-6215-15-173
Cite this article as: Nørskov *et al.*: Detailed statistical analysis plan for the difficult airway management (DIFFICAIR) trial. *Trials* 2014 **15**:173.

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.

Nørskov AK[1, 2], Wetterslev J[2], Rosenstock CV[1], Afshari A[3], Astrup G[4], Jakobsen JC[2], Thomsen JL[5], Bøttger M[6], Ellekvist M[3], Schousboe BMB[7], Horn A[9], Jørgensen BG[10], Lorentzen K[11], Madsen MH[1], Knudsen JS[12], Thisted BK[13], Estrup S[14], Mieritz HB[15], Klesse T[16], Martinussen HJ[17], Vedel AG[18], Maaløe R[19], Bøsling KB[20], Kirkegaard PRC[21], Ibáñez CR[22], Aleksandraviciute G[23], Hansen LS[24], Mantoni T[25], Lundstrøm LH[1]

[1]Department of Anesthesiology, Copenhagen University Hospital, Nordsjællands Hospital, Hillerød, Denmark (work should be attributed to this department); [2]Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [3]Department of Anesthesiology, Juliane Marie Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [4]Department of Anesthesiology, Aarhus University Hospital, Aarhus, Denmark; [5]Department of Anesthesiology, Copenhagen University Hospital, Herlev Hospital, Herlev, Denmark; [6]Department of Anesthesiology, Centre of Head and Orthopedics, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [7]Department of Anesthesiology, Neuroscience Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [9]Department of Anesthesiology, Bekkevold Private Hospital, Hellerup, Denmark; [10]Department of Anesthesiology, Frederiksberg Hospital, Copenhagen University Hospital, Frederiksberg, Denmark; [11]Department of Anesthesiology, Glostrup Hospital, Copenhagen University Hospital, Glostrup, Denmark; [12]Department of Anesthesiology, Lillebælt Hospital, Kolding, Denmark; [13]Department of Anesthesiology, Copenhagen Private Hospital, Copenhagen, Denmark; [14]Department of Anesthesiology, Nykøbing Falster Hospital, Copenhagen University Hospital, Nykøbing Falster, Denmark; [15]Department of Anesthesiology, South Jutland Hospital, Aabenraa, Denmark; [16]Department of Anesthesiology, South Jutland Hospital, Sønderborg, Denmark; [17]Department of Anesthesiology, Private Hospital Kollund, Kollund, Denmark; [18]Department of Anesthesiology, Heart Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark; [19]Department of

Anesthesiology, Bispebjerg Hospital, Copenhagen University Hospital, Copenhagen, Denmark [20]Department of Anesthesiology, Roskilde Hospital, Copenhagen University Hospital, Roskilde, Denmark; [21]Department of Anesthesiology, Næstved Hospital, Copenhagen University Hospital, Næstved, Denmark; [22]Department of Anesthesiology, South Jutland Hospital, Haderslev, Denmark [23]Department of Anesthesiology, Thy-Mors Hospital, Thisted, Denmark; [24]Department of Anesthesiology, Lillebælt Hospital, Vejle, Denmark; [25]Department of Anesthesiology, Abdominal Centre, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark
[*]Corresponding author: anderskehlet@hotmail.com

Acknowledgements

The Danish Anaesthesia Database and the participating departments.

Funding

The trial is funded by independent grants by: The TRYG foundation; Nordsjællands Hospitals research council; The National Research Council of Denmark; The Capital Regions funds for research; DASAIMs fund.

Word count

Abstract: 249

Introduction: 499

Discussion: 1,317

Abbreviated title

Cluster randomized trial comparing airway assessments

Abstract

Background

Unanticipated difficult intubation remains a pivotal challenge in anesthesia. The Simplified Airway Risk Index (SARI) is a risk model for prediction of difficult laryngoscopy.

Our aim was to compare airway assessment based on the SARI with usual care.

Methods

From 01.10.2012 to 31.12.2013, 26 departments were cluster-randomized to either apply the SARI model or usual care for airway assessment. The pre-operative prediction of difficult intubation was registered in the Danish Anaesthesia Database. Actual intubation difficulties were recorded. The primary outcomes were the proportions of unanticipated difficult and unanticipated easy intubation.

Results

In the primary analyses 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 was 1.03 (95% CI: 0.77–1.38), P=0.84. The proportion of unanticipated easy intubation was 1.42% (415) in SARI departments versus 1.00% (302) in usual care departments. Adjusted OR was 1.26 (0.68–2.34), P=0.47.

In SARI departments neither a 58% increase in patients anticipated to have intubation difficulties nor an 84% increase in patients scheduled for advanced intubation techniques reached adjusted statistical significance, P = 0.29 and P = 0.06 respectively.

Conclusions

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, it may increase the anticipation of intubation difficulties and change practice towards advanced intubation techniques.

Introduction

Prediction of difficult airway management remains a pivotal challenge in anesthesia. Unanticipated difficult tracheal intubation can cause serious patient complications and identification of at risk patients is highly prioritized among anesthesia personnel. Accurate prediction of difficult intubation may reduce potential complications by the allocation of experienced personnel and by using relevant equipment [1]. No single predictor is sufficiently valid in predicting difficult intubation [2–6]. However, several studies indicate that combining multiple predictors of difficult intubation will increase the predictive value of the assessment [6]. All major anesthesia societies recommend pre-operative airway assessment[7–9]. Despite a general agreement about the need and rationale for a pre-operative airway assessment, it still remains unclear how this assessment should be performed. The American Society of Anesthesiologists (ASA) recommends a pre-operative airway assessment based on eleven anatomical variables. However, they do not elaborate on which factors are considered mandatory for examination nor on how they should be weighted in an overall assessment, but argues that it depends on the clinical context[8]. Likewise, in the UK, the NAP4 gives no elaboration on the content of the airway assessment[10]. Consequently the choice of airway assessment is left to the discretion of the individual anesthesiologist and prediction of difficult intubation is based on the response to the question: do I anticipate a difficult intubation? [11]. This anticipation may or may not be based on a wide array of pre-operative airway examinations, depending on the individual anesthesiologist and various department recommendations. Pre-operative airway assessment is widely implemented in Denmark and we find it fair to assume that it is based on patient examinations for one or several known predictors of difficult intubation [12]. Previously, we have reported the diagnostic accuracy of the individual anesthesiologist to be low to moderate, with 75 to 93% of alle difficult intubations being unanticipated difficult[11]. The "Simplified Airway Risk Index" (SARI) is a multivariable model for airway assessment [13]. It enables an estimation of the likelihood of a difficult

direct laryngoscopy. The clinical impact of most predictive models are not assessed before being introduced into a clinical setting[14]. This induces risks of over-estimating the predictive potential of the model[14, 15]. Assessment of the impact of introducing a new model requires a comparative design and is optimally done in a (cluster) randomized trial, compared with usual care[15, 14]. No randomized clinical trial has compared using a predictive model for airway difficulties with usual care. We aimed to compare the SARI as a systematic airway assessment tool with usual care on the proportion of unanticipated difficult intubation. We hypothesized that a systematic use and registration of the SARI, combined with continuous education, could reduce the relative risk of an unanticipated difficult intubation with 30% compared with usual care. We considered a relative risk reduction of 30% to be clinically important. Based on data from 2011, detection of such an intervention effect would correspond to a number needed to treat (NNT) of 180. Meaning, prevention of one unanticipated difficult intubation for every 180 airway assessed.

Methods

Trial design

The DIFFICAIR trial is a cluster randomized trial, stratified according to the proportion of unanticipated difficult intubation at department level in 2011. A total of 28 Danish departments of anesthesia were randomized 1:1 to SARI airway assessment (SARI departments) or usual care (usual care departments). The trial was conducted from 01.10.2012 to 31.12.2013. Assessing the impact of a new strategy implementation, on a departmental level, makes individual patient randomization impossible, since the anesthesiologists would face difficulties in distinguishing different methods for airway assessment for different patients. Thus, a clustered trial design was required. Prior to data extraction, a trial protocol describing the trial design, as well as a detailed statistical analysis plan, were published and made available on http://www.clinicaltrials.gov (NCT01718561) [16, 17].

SARI departments

In the intervention group (SARI departments) anesthesiologists were repeatedly trained in pre-operative use of the SARI score in order to ensure uniform and high quality airway assessments [18]. All adult patients were planned to be airway assessed, using the following fixed panel of predetermined predictors for difficult intubation: 1) mouth opening, 2) thyromental distance, 3) modified Mallampati classification, 4) neck movement, 5) ability to prognath, 6) weight and 7) history of difficult intubation. Each predictor represents a weighted score of 0-1 or 0-2 points. A summarized score (the SARI score) of ≥ 4 was indicative of difficult direct laryngoscopy [13] *(Figure 1)*.The original SARI uses the original Mallampati grade, whereas a modified Mallampati class was used in the DIFFICAIR trial[16, 19]. To help implement and sustain the intervention, continuous education in correct use of the SARI was conducted throughout the trial period, and an array of tutorial aids (a video, posters, white coat aids etc.) was produced and distributed before trial initiation *(Figure 2)*.

Usual care departments

The departments in the control group continued their existing recommendation on pre-operative airway assessment. All potential airway-related variables were recorded as previously on the anesthesia record.

Data registration in the Danish Anaesthesia Database

Assessment of all outcomes was based on data recorded in the DAD. The database contains quantifiable indicators, covering the peri-operative period. Regardless of trial group, all anesthesiologists had to tick a Y/N box to answer a mandatory question regarding anticipation of difficult intubation by direct laryngoscopy. Furthermore, a scheduled airway management plan was recorded pre-operatively. Immediately following the airway management an intubation score was registered *(Figure 3)*. The usual care departments continued to record what was already implemented and mandatory in the DAD. The SARI departments recorded a more elaborate pre-operative airway assessment consisting of 'the individual variables included in the SARI model' and 'the SARI score'. The anesthesiologist's anticipation of intubation difficulties (Y/N) was also recorded on the SARI departments. The DAD auto-generated a SARI score based on the variables registered in the SARI model. The SARI score was merely meant to guide and supplement the prediction of intubation difficulties. Thus, the anesthesiologist's prediction of intubation difficulties could differ from the prediction contained in the SARI score. The outcome measures were based on the anesthesiologist's predictions and not the SARI score, hence being comparable between trial groups.

Definition of difficult intubation

An intubation score based on the numbers of intubation attempts and the use of equipment is an integrated part of the DAD while adhering to the Canadian Airway Focus Group definition of difficult

intubation[9]*(Figure 3)*. Thus, the score used in the DIFFICAIR trial represents the conditions in relation to intubation instead of the mere laryngoscopic view used as a surrogate measure for difficult intubation defined by El-Ganzouri et al[13, 20]. We found it clinically more relevant to use an intubation score based on presence or absence of actual intubation difficulties.

Data monitoring

A local investigator was appointed at all departments. During the trial period, the degree of data completeness was monitored on a monthly basis for each department. In case of a declining registration of the SARI, the local investigator was approached in order to restore the registration rate. Data were retrieved from the DAD, securing patient anonymity in accordance to regulations by the Danish Data Protection Agency.

Participants

Eligibility on cluster level

All departments that were expected to include a minimum of 200 intubated patients in the trial period based on their previous data recording in the DAD during 2011.

Eligibility on patient level

Adult patients (≥15 years) that were attempted intubated, regardless of scheduled airway management plan. Patients registered with more than one episode of surgery and airway management during the trial period were included with their first entry [21].

Outcomes

All outcomes were measured on an individual participant level.

Primary outcome 1 was ‘the proportion of participants with an unanticipated difficult intubation’. Primary outcome 2 was ‘the proportion of participants with an unanticipated easy intubation’. Secondary outcome 1 was 48 hours mortality. Secondary outcome 2 was 30 days mortality. Secondary outcome 3 was ‘the proportion of participants anticipated difficult to intubate, scheduled for and attempted intubated by an advanced method’. Secondary outcome 4 was 1-sensitivity. Secondary outcome 5 was sensitivity. Secondary outcome 6 was specificity. Secondary outcome 7 was positive predictive value. Secondary outcome 8 was negative predictive value. Secondary outcome 9 was positive likelihood ratio. Secondary outcome 10 was negative likelihood ratio.

Primary analyses *(Figure 4)*

Population 1: Patients that were attempted intubated but not pre-operatively scheduled for advanced intubation methods

According to the intubation score *(Figure 3)*, patients were categorized as difficult to intubate if they were intubated using a more advanced technique than direct laryngoscopy (e.g. video laryngoscopic intubation), regardless of the reason behind this choice (e.g. educational purposes).

Therefore, we excluded patients who were pre-operatively scheduled for intubation by advanced techniques. If a patient was intubated with an advanced technique despite being scheduled for conventional laryngoscopy, we assumed that the change of intubation equipment was due to difficulties with direct laryngoscopy, thus representing a difficult intubation.

Population 2: Patients that were attempted intubated but not pre-operatively scheduled for advanced intubation methods plus patients anticipated difficult to intubate, scheduled for, and attempted intubated by an advanced method

In order to avoid an erroneous exclusion of correctly identified difficult intubations, we defined population 2. We assumed that if the anesthesiologist had predicted a difficult intubation, this was indeed the reason for choosing an advanced technique. Consequently, we identified a group of patients

who were predicted difficult to intubate, scheduled for and intubated by an advanced intubation technique. Accordingly, the assumption for population 2 was that these patients were correctly identified as difficult to intubate, and they were included as such in addition to the patients from population 1.

Sensitivity analyses *(Figure 4)*

Sensitivity analysis 1

To explore whether the accuracy of the predictions improves with increasing severity of intubation difficulties, we performed a sensitivity analysis choosing a more rigorous definition of difficult intubation (intubation score ≥ 3) than the predefined definition in DAD (intubation score ≥ 2) *(Figure 3).*

Sensitivity analysis 2

We performed a sensitivity analysis comparing the SARI model's predictive accuracy from patients in the SARI departments with the anesthesiologists' predictions of difficult intubation in the usual care departments. We defined a SARI score ≥ 4 as anticipation of difficult intubation in SARI departments.

Randomization and blinding

According to a computer-generated allocation sequence, the departments were randomly assigned to SARI or usual care. The departments were stratified into two strata, based on the proportion of patients registered as unanticipated difficult to intubate, from data registered during 2011 in DAD, the strata being < 2% or ≥ 2%.

With appropriate use of allocation concealment, all the heads of departments provided written informed consent to trial participation prior to randomization of the departments.

The trial design made it impossible to blind the patients and the participating anesthesiologists. The authors were blinded for the allocation of departments to SARI- or usual care when performing the

initial analyses on the primary outcomes. Two different blinded manuscripts were written considering the SARI departments being in one or the other of the two groups. After having agreed on both manuscripts, including conclusions, the steering committee unveiled the blinding, and the corresponding conclusion was employed [22].

Statistics

Sample size

Sample size estimation was performed prior to randomization using proper adjustment for the cluster randomized trial design. We adjusted the required individual sample size for the intra cluster correlation by calculating the between and within cluster variances on 2011 data for the included clusters [23].

A minimum of 26 departments with an average cluster size of approximately 2,500 patients was required to detect or reject a relative risk reduction of 30% assuming a usual care event proportion of the primary outcome of 1.87%, a risk of type I error of 5% and a risk of type II error of 20%.

Comparisons between the outcomes of the trial groups were done on an individual participant level according to our sample size estimation. Results were presented by frequencies, proportions, percentages, odds ratios, risk ratios, and 95% confidence intervals (CI). The primary analyses of the primary outcomes were done according to the ITT principle and adjusted for design variables (the stratification- and the cluster variable) using generalized estimating equations (GEE). The GEE analysis method was used to account for within cluster correlation [24–26].

We tested the robustness of the results by repeating the analyses in a mixed effects model and with a standard t-test on means at departmental level. Further, we tested the primary outcome using GEE adjusting for design variables and the pre-defined assumed confounders age; gender; ASA classification; emergency/elective procedure; body mass index (BMI); and use of neuromuscular blocking agents.

A chi-squared test was used to assess the effect of the intervention on binary secondary outcomes. For the adjusted analyses, GEE were performed.

Handling of incomplete data

We used multiple imputation for handling of missing data exceeding a rate of 5% and with a statistical significant Little's test precluding analyses of complete cases[27–30]. The imputation was made assuming data missing at random. We imputed the dichotomised variable SARI score less than or ≥ 4 and only for patients in SARI departments having missing values. Using logistic regression, we imputed from patients within the same cluster to account for between cluster variance. The following variables were used: age; BMI; unanticipated difficult intubation; SARI score less than or ≥ 4. Based on the imputed variable we constructed the outcome variable 'difficult intubation and SARI < 4'. We did 10 imputations and compared them with the usual care departments in a cumulated GEE using STATA version 13.

Trial registration and Ethics

The Committee on Health Research Ethics of the Capital Region of Denmark declared that the DIFFICAIR trial was regarded as a quality assurance project without risk, side effects or inconvenience for the patients, as the trial protocol included no specific dictation on airway management. Thus, the trial was exempted from the committee system (H-3-2012-FSP2). Further, the need for individual patient consent was waived. However, informed consent from every participating department by the Head of Department was acquired before randomization. The trial was approved by The Danish Data Protection Agency (2007-58- 0015/HIH-2011-10, I-Suite nr: 02079) and is registered at http://www.clinicaltrials.gov (NCT01718561). Reporting of the trial was done according to the CONSORT statement: extension to cluster randomized trials [31].

Results

Two usual care departments never initiated DAD registration and were excluded giving a total number of 26 clusters (15 SARI clusters and 11 usual care clusters *(Figure 4)*).

We retrieved a total of 75,799 entries of intubation attempts. The SARI departments registered 37,801 and the usual care departments 37,998 patients. When including the first entry for each patient 32,358 patients were attempted intubated in SARI departments and 31,915 patients in usual care departments. In SARI departments, 78.4% had a registration sufficient enough to classify the SARI score as ≥4 or below 4. Hence, 21.6% of the patients in these departments did not have a SARI registration sufficient enough to classify the score as ≥4 or below 4.

Baseline characteristics of patients and clusters are shown in Tables 1 and 2.

Primary outcomes:

Population 1 (patients that were attempted intubated but not pre-operatively scheduled for advanced intubation methods)

Primary outcomes were measured on 59,514 patients not scheduled for advanced intubation *(Figure 4)*. There were 29,209 patients in SARI departments and 30,305 the usual care departments. The incidence of difficult intubation was 2.66% (778) in SARI departments and 2.62% (794) in usual care departments.

There were 2.38% (696) unanticipated difficult intubations in SARI departments and 2.39% (723) in the usual care departments. OR adjusted for stratification and cluster in a GEE model was 1.03 (0.77–1.38), P=0.84. By further adjusting for the assumed confounders age; gender; ASA classification; emergency/elective procedure; BMI; and use of neuromuscular blocking agents in a GEE the OR was 1.17 (0.81-1.67), P=0.40.

The proportion of unanticipated easy intubation was 1.42% (415) in the SARI versus 1.00% (302) in usual care departments. Adjusted OR was 1.26 (0.68–2.34), P=0.47. Mixed model analysis and student's t-test yielded similar results.

Population 2 (patients that were attempted intubated but not pre-operatively scheduled for advanced intubation methods plus patients anticipated difficult to intubate, scheduled for, and attempted intubated by an advanced method)

This analysis included 60,609 patients (29,934 in the SARI versus 30,675 in usual care departments). The proportion of unanticipated difficult intubations in SARI departments was 2.32% (694) versus 2.36% (723) in usual care departments. Adjusted OR was 1.02 (0.77-1.35), P=0.89.

Sensitivity analysis 1 (all patients included and difficult intubation defined as an intubation score ≥ 3)

In this analysis, we applied a more rigorous definition of difficult intubation including 32,358 in the SARI and 31,915 in usual care departments. In SARI departments 1.16% (375) was unanticipated difficult to intubate in comparison with 1.21% (385) in usual care departments. Adjusted OR was 1.01 (0.73-1.39), P=0.96.

Sensitivity analysis 2 (the diagnostic accuracy of the SARI score compared with usual care departments)

We used a SARI score ≥ 4 as the definition of anticipation of difficult intubation. In a per protocol analysis of the patients having a sufficient SARI score (n=22,826) the proportion of unanticipated difficult intubation was 2.17% (496) in SARI departments compared with 2.39% (723) in usual care departments. Adjusted OR was 0.95 (0.70-1.28), P=0.72. Using multiple imputation to handle missing data on the SARI score, the adjusted OR was 0.93 (0.68-1.26), P=0.63.

Secondary outcomes:

Mortality

We found no statistically significant difference between the two groups on 48 hour or 30 day mortality. Adjusted OR was 1.27 (0.63-2.59), P=0.51 for 48 hour mortality and 0.69 (0.37-1.78), P=0.24 for 30 days mortality.

The proportion of participants anticipated difficult to intubate, scheduled for and attempted intubated by an advanced method

The number of patients anticipated difficult to intubate, scheduled for and intubated by an advanced method was 2.21% (714) in SARI departments versus 1.18% (378) in usual care departments. Adjusted OR was 1.31 (0.54-3.17), P=0.55.

Diagnostic accuracy

Of all difficult intubations in population 1, the fraction of unanticipated difficult intubation was 89.5 % (696/778) in SARI departments and 91.1 % (723/794) in usual care departments; adjusted OR = 1.02 (0.59-1.78), P=0.93. Further, no adjusted analyses yielded statistical significant differences in sensitivity, specificity, positive and negative predictive values or positive and negative likelihood ratios between the trial groups.

Scheduled airway management

In SARI departments there was an 84% increase in the proportion of patients scheduled for an advanced intubation methods compared with usual care departments, 10.33% (3,342/32,358) versus 5.62% (1,794/31,915). Adjusted OR was 2.50 (0.98-6.37), P=0.06.

Concordantly, we found a 58% increase in the proportion of patients anticipated difficult to intubate in SARI departments compared with usual care departments (4.32% (1,397/32,358) versus 2.73% (871/31,915)). Adjusted OR was 1.35 (0.77-2.38), P=0.29.

Discussion

Education in the use of the SARI, implementation of the SARI as a new guideline and its registration in DAD did not lead to a significant reduction in the proportion of unanticipated difficult and easy intubation. Our conclusions remain similar regardless of the choice of population or performed sensitivity analyses. If the intervention did improve the diagnostic accuracy this should consequently lead to a lower number of unanticipated difficult intubations in SARI departments. While this was not the case, the behavior of the anesthesiologists did tend to change in SARI departments, leading to an increase in the point estimates of the number of patients predicted difficult to intubate and in the number of patients scheduled for advanced intubation techniques.

Prior to the DIFFICAIR trial we conducted a baseline cohort study on DAD data from 2008-2011[11]. The frequency of difficult intubation was 1.86% in this study. In the DIFFICAIR trial the frequency of difficult intubation was 2.66 % and 2.62 % (SARI versus usual care). The enhanced focus on correct registration during the DIFFICAIR trial might explain the increased frequency of difficult intubation compared to baseline registration, rather than an actual increase in intubation difficulties. The increased frequency of intubation difficulties enhanced the power of our trial compared to the estimated sample size calculations that were based on data from the DAD in 2011. Prior observational studies on prediction of difficult intubation have been conducted under rigorous study settings and have demonstrated average to good predictive accuracy of the examined models. The DIFFICAIR trial on the other hand is a cluster randomized clinical trial comparing the clinical impact of implementing a systematic strategy for pre-operative airway assessment in daily clinical practice with usual care[14, 15]. Furthermore, data reflects daily practice from a widespread population of surgical patients and a heterogeneous group of anesthesia providers, and thus outcome assessers. Comparison on prediction rates from this RCT with prior observational studies should be made with caution.

Our trial has a number of strengths: 1) We applied state of the art methodology for testing the clinical impact of a predictive model [14, 15, 32], 2) the trial methodology was prospectively planned and reported in a published protocol and a statistical analysis plan [16, 17], 3) the applied methodology reduced the risk of systematic error ('bias') [33], 4) the large number of participants reduced the risk of random error ('play of chance') [34], and 5) adequate statistical methods were used to account for the clustered nature of the data (GEE) and for missing data (multiple imputation). 5) In order to explore a potential baseline imbalance, we performed a post hoc cluster adjusted GEE analysis, analogue to the one used on the present data, on 2011 baseline data comparing the departments later randomized to SARI departments with departments later randomized to usual care departments. We found no baseline difference between the two groups on the proportion of unanticipated difficult intubation, implicating no baseline imbalance on the primary outcome before trial initiation (OR 0.98, (95% CI: 0.72-1.35), P=0.88).

Our trial design has a number of limitations. 1) Since the outcome assessor could not be blinded, the same person conducting the pre-operative assessment could theoretically also perform the outcome assessment. It was not possible to externally validate the data registered in the DAD, since no other registry records these data in Denmark. Most departments do follow up registration on missing patients, but we cannot rule out, that some undergoing anesthesia were never registered. 2) A local investigator was appointed on each department. Prior to trial commencement, the personnel on interventional departments were educated in using the SARI and in how to register airway assessment and management correctly in the DAD. Local investigators in the usual care departments were required to teach correct registration in the DAD and written instructions were distributed to all relevant personnel. The different level of attention and education may have led to an increased awareness and registration of difficult intubations in the SARI departments.

3) In the design of the trial we sought to minimize the risk of contamination bias from the SARI to the usual care departments. But we could not prohibit the anesthesiologists in the usual care departments

from using the SARI score. Further, some anesthesiologists have unarguably changed employment, moving from SARI departments to usual care departments, potentially inducing minor contamination bias. It was impossible to conduct the trial unnoticed in Denmark. Therefore, a change of behavior towards using airway assessment resembling the SARI departments' might have happened as a spillover effect on usual care departments. Additionally, some patients were impossible to SARI assess and some anesthesiologists unquestionably either forgot or deliberately avoided the use of the SARI. The fact that not all patients in the SARI departments were completely assessed by the SARI may have obscured a true intervention effect. Further, we could not ethically dictate the anesthesiologists to abide by the predictions of the SARI score and we found a discrepancy between the individual predictions of the anesthesiologists and the cut off value for anticipating a difficult intubation comprised in the SARI (SARI $\geq$ 4). However, in our analyses using this cut off value for distinguishing the anticipation of difficult intubation, we found no statistical significant difference on the primary outcome.

4) The cluster randomized trial is inherently more prone to risk of baseline imbalance than the individually randomized trial and results should be interpreted baring this in mind. Even though the number of patients in each trial group was almost perfectly balanced, the case-mix was slightly uneven. The patients in usual care departments were slightly older and with a tendency of having a higher ASA classification. On the other hand more patients in SARI departments were intubated without the use of neuromuscular blocking agents. These variables are all potentially associated with difficult intubation[35]. Therefore adjusted analyses were pre-planned in the protocol and performed accordingly. However, we cannot rule out potentially unknown residual confounding.

No other randomized clinical trial has prospectively compared two different strategies for pre-operative airway assessment. Although the trial was completed successfully from both a methodological and practical point of view, our results may not be easily reproduced. The trial settings

involved a national clinical database and 26 departments accepting to implement a new guideline for pre-operative airway assessment before randomization.

The intervention in the DIFFICAIR trial was a combination of systematically applying the SARI for all patients, and a thorough education of physicians and nurses in using this tool. Based on this training, and the use of the tool, we hypothesized that the anesthesiologists would improve their ability to predict difficult intubation. In this trial the 'true' accuracy of the anesthesiologists' predictions of intubation difficulties probably lies somewhere between the predictive values found in population 1 and 2; with population 1 representing a 'worst case' scenario and population 2 representing a 'best case' scenario. Our results do not support that a pre-operative use of the SARI, leads to a reduction in the incidence of unanticipated difficult intubations.

The main aim of predicting difficult intubations is to avoid airway management related morbidity, ranging from simple tooth injuries to anoxic brain damage or even death. The SARI departments seemingly categorized more patients at risk of intubation difficulties and allocated a larger number of patients to advanced intubation methods. Allocation to advanced intubation methods was undoubtedly the correct approach for some patients. However, it may equally have been superfluous for others, resulting in unnecessary use of resources and potential patient discomfort [36]. It is debatable, which ratio of false versus true predictions is to be considered acceptable, and our data do not allow us to exactly estimate such ratio [14, 37].

Based on the present trial we cannot recommend the SARI model as a superior approach to pre-operative airway assessment compared with usual care. The proportions of unanticipated difficult intubation found in the DIFFICAIR trial underline the continued challenge anesthesiologists' face in predicting these events.

Conclusion

Implementation of a systematic use and registration of the SARI as a strategy for pre-operative airway assessment could not be demonstrated to reduce the incidence of unanticipated difficult or easy intubation. In SARI departments a larger number of patients may have been anticipated difficult to intubate, and a change in practice towards using more advanced intubation techniques may have occurred.

Abbreviations

ASA: American Society of Anesthesiologists

UK: United Kingdom

NAP4: Fourth National Audit Project

DAD: Danish Anaesthesia Database

SARI: Simplified Airway Risk Index

NNT: Number Needed to Treat

BMI: Body Mass Index

RCT: Randomized Clinical Trial

ITT: Intention to Treat

OR: Odds Ratio

GEE: Generalized Estimating Equation

ENT: Ear Nose Throat

Declaration of interests

The authors declare that they have no competing interests.

Authors' contributions

Steering Committee of the DIFFICAIR trial: Nørskov, Wetterslev, Rosenstock, Astrup, Afshari, Lundstrøm

Study concept and design: Nørskov, Wetterslev, Rosenstock, Lundstrøm

Acquisition of data: Nørskov, Wetterslev, Rosenstock, Afshari, Thomsen, Bøttger, Ellekvist, Mantoni, Schousboe, Vedel, Horn, Jørgensen, Lorentzen, Madsen, Knudsen, Thisted, Estrup, Mieritz, Klesse, Maaløe, Bøsling, Kirkegaard, Ibanez, Alexandraviciute, Hansen, Martinussen, Lundstrøm

Analyses and interpretation of data: Nørskov, Wetterslev, Rosenstock, Astrup, Afshari, Jakobsen, Lundstrøm

Drafting of the manuscript: Nørskov, Wetterslev, Rosenstock, Astrup, Afshari, Lundstrøm

Critical revision of the manuscript for important intellectual content: Nørskov, Wetterslev, Rosenstock, Astrup, Afshari, Jakobsen, Thomsen, Bøttger, Ellekvist, Mantoni, Schousboe, Vedel, Horn, Jørgensen, Lorentzen, Madsen, Knudsen, Thisted, Estrup, Mieritz, Klesse, Maaløe, Bøsling, Kirkegaard, Ibanez, Alexandraviciute, Hansen, Martinussen,, Lundstrøm

Statistical analyses: Nørskov, Wetterslev, Jakobsen, Lundstrøm

Obtained funding: Nørskov, Wetterslev, Rosenstock, Lundstrøm

References

1. Cook TM, Macdougall-Davis SR: **Complications and failure of airway management**. *Br J Anaesth* 2012, **109**(Supplement 1):i68–i85.

2. 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.

3. 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.

4. 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.

5. 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.

6. Shiga T, Wajima Z, Inoue T, Sakamoto S, Sakamoto A: **Predicting difficult intubation in apparently normal patients**. *Anesthesiology* 2005, **103**:429–437.

7. 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.

8. 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.

9. 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.

10. 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.

11. Nørskov a K, 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.

12. Nørskov AK, Rosenstock CV, Wetterslev J, Lundstrøm LH: **Pre-operative airway assessment. No national consensus in Denmark.** *Eur J Anaesthesiol* 2014, **31**(e-suppl 52, Abstract programme):275.

13. El-Ganzouri AR, McCarthy RJ, Tuman KJ, Tanck EN, Ivankovich AD: **Pre-operative airway assessment: predictive value of a multivariate risk index**. *Anesth Analg* 1996, **82**:1197–1204.

14. Moons KGM, Altman DG, Vergouwe Y, Royston P: **Prognosis and prognostic research: application and impact of prognostic models in clinical practice.** *BMJ* 2009, **338**(June 2009):b606.

15. Steyerberg EW, Moons KGM, Windt DA Van Der, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG, Group P: **Prognosis Research Strategy (PROGRESS) 3: prognostic model research**. *PLoS Med* 2013(February).

16. 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.

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

18. Rosenstock C, Ostergaard D, Kristensen MS, Lippert A, Ruhnau B, Rasmussen LS: **Residents lack knowledge and practical skills in handling the difficult airway**. *Acta AnaesthesiolScand* 2004, **48**:1014–1018.

19. Samsoon GLT, Young JRB: **Difficult tracheal intubation: a retrospective study.** *Anaesthesia* 1987, **42**:487–490.

20. Cormack RS, Lehane J: **Difficult tracheal intubation in obstetrics.** *Anaesthesia* 1984, **39**:1105–11.

21. Altman DG, Bland JM: **Statistics notes. Units of analysis.** *BMJ* 1997, **314**:1874.

22. Gøtzsche PC: **Blinding during data analysis and writing of manuscripts.** *Control Clin Trials* 1996, **17**:285–90.

23. Hayes RJ, Bennett S: **Simple sample size calculation for cluster-randomized trials.** *Int J Epidemiol* 1999, **28**:319–26.

24. Kahan B, Morris T: **Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis**. *BMJ* 2012, **345**:e5840.

25. Klar N, Donner A: **Current and future challenges in the design and analysis of cluster randomization trials**. *Stat Med* 2001, **20**:3729–3740.

26. Kahan BC, Morris TP: **Assessing potential sources of clustering in individually randomised trials.** *BMC Med Res Methodol* 2013, **13**:58.

27. **Multiple Imputation for Nonresponse in Surveys - Rubin - Wiley Online Library** [http://onlinelibrary.wiley.com/book/10.1002/9780470316696]

28. Schafer JL: **Multiple imputation: a primer.** *Stat Methods Med Res* 1999, **8**:3–15.

29. Schafer JL, Graham JW: **Missing data: our view of the state of the art.** *Psychol Methods* 2002, **7**:147–77.

30. Hickam D, Totten A, Berg A, Rader K, Goodman S, Newhouse R, PCORI Methodology Committee: **The PCORI Methodology Report 2013**. *http://www.pcori.org/assets/2013/11/PCORI-Methodology-Report.pdf* .

31. Campbell MK, Piaggio G, Elbourne DR, Altman DG: **Consort 2010 statement: extension to cluster randomised trials.** *BMJ* 2012, **345**:e5661.

32. Riley RD, Hayden JA, Steyerberg EW, Moons KGM, Abrams K, Briggs A, Schroter S, Altman DG, Kyzas PA: **Prognosis Research Strategy (PROGRESS) 2: prognostic factor research**. *PLoS Med* 2013(February).

33. Higgins JPT, Green S: *Cochrane Handbook for Systematic Reviews of Interventions*. *Volume 5.0.1*; 2008:672.

34. 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.

35. Lundstrøm LH, Møller AM, Rosenstock C, Astrup G, Gätke MR, Wetterslev J: **Avoidance of neuromuscular blocking agents may increase the risk of difficult tracheal intubation: a cohort study of 103,812 consecutive adult patients recorded in the Danish Anaesthesia Database.** *Br J Anaesth* 2009, **103**:283–90.

36. Yentis SM: **Predicting difficult intubation--worthwhile exercise or pointless ritual?**. *Anaesthesia* 2002, **57**:105–9.

37. O'Sullivan E, Laffey J, Pandit JJ: **A rude awakening after our fourth "NAP": lessons for airway management**. *Anaesthesia* 2011, **66**(1365-2044 (Electronic)):331–334.

Figure legends

Figure 1:

Pre-operative airway assessment for each trial group

Figure 2:

White coat aid

Figure 3:

Data registered in the Danish Anaesthesia Database

Figure 4:

Flow diagram of cluster and patient allocation

Table legends

Table 1:

Cluster level summaries

Table 2:

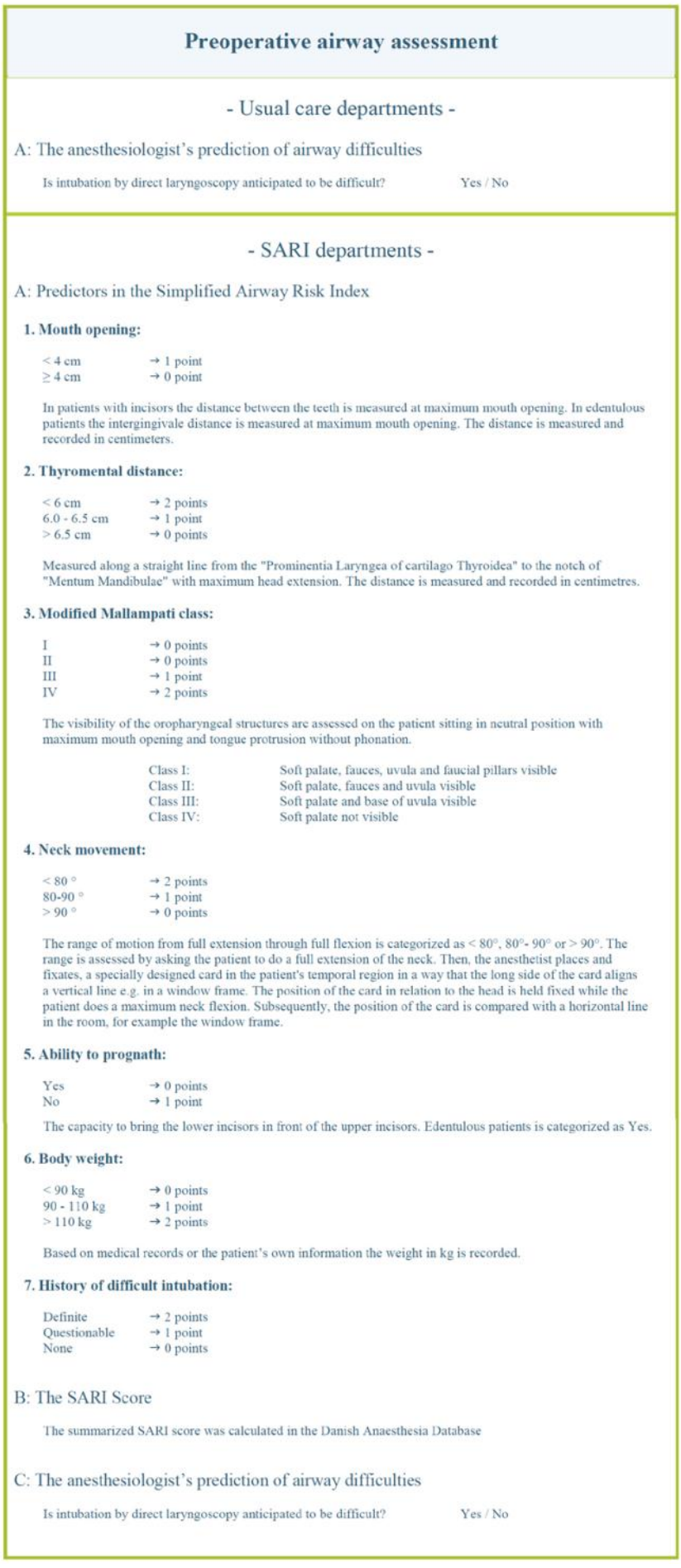
Individual patient level summaries

Table 3:

Diagnostic accuracy for different populations and sensitivity analyses

Figure 1

Pre-operative airway assessment for each trial group



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

Figure 2

White coat aid

Figure 3

Data registered in the Danish Anaesthesia Database

Preoperative airway assessment

The anesthesiologist's prediction of airway difficulties

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

Scheduled airway management plan

For each patient one of the following options is chosen:

1. None / unknown
2. Spontaneous breathing
3. Mask ventilation
4. Laryngeal mask (any kind)
5. Intubation via direct laryngoscopy
6. Intubation via video laryngoscope
7. Intubation via flexible fiber-optic scope
8. Intubation via another method (e.g. Fastrach)
9. Tracheotomy under local anesthesia
10. Already intubated or tracheotomized

Actual intubation conditions

Intubation

Intubation is graded according to the following score. One of the below options is chosen in succession of the airway management procedure:

0. Not attempted
1. Maximum two intubation attempts – Only by direct laryngoscopy
2. Maximum two intubation attempts in which other intubation equipment (e.g. video laryngoscope) is used
3. Three intubation attempts or more - Regardless of intubation method
4. Intubation failed despite attempting

Tracheal intubation by direct laryngoscopy is defined as unproblematic by a score = 1 and difficult by a score ≥ 2

Figure 4

Flow diagram of cluster and patient allocation

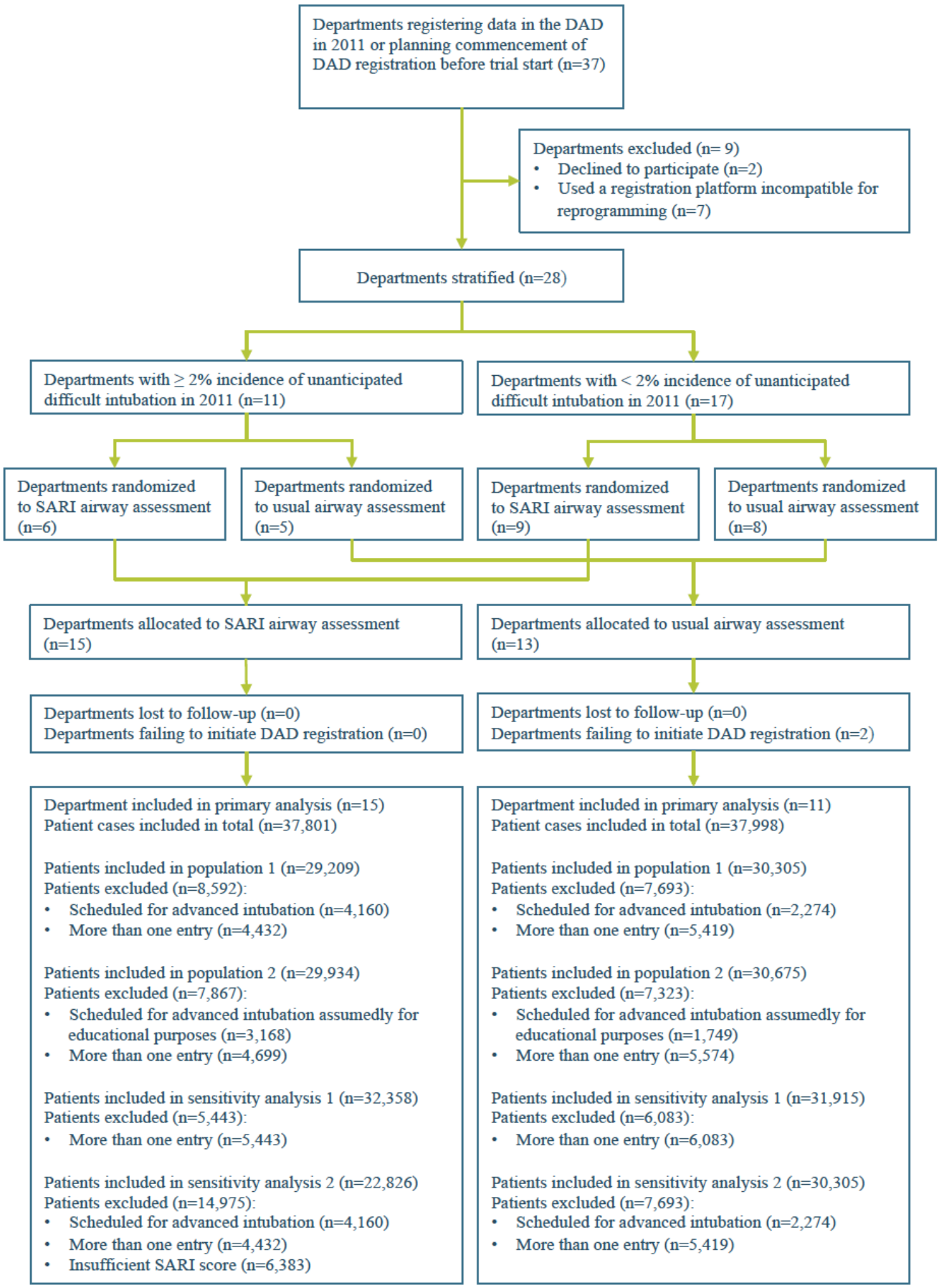


Table 1

Cluster level summaries

	SARI departments (15 clusters)	**Usual care departments** (11 clusters)
Mean cluster size (All patients in general anesthesia)	4977	6900
Mean number of intubated patients	2157	2901
Mean fraction of unanticipated difficult intubations in per cent (Population 1)	2.02	1.93
Mean fraction of unanticipated easy intubations in per cent (Population 1)	1.85	1.05
Mean Age	53.2	54.4
Mean BMI	26.4	26.4
Mean fraction ASA I	41.4	35.3
Mean fraction ASA II	45.1	42.3
Mean fraction ASA III	11.3	19.5
Mean fraction ASA IV	1.0	2.5
Mean fraction ASA V	0.1	0.1
Mean fraction Unknown ASA class	1.3	0.3
Fraction of private hospitals	0.27	0.18
Fraction of stratum 'high' (≥ 2% unanticipated difficult intubations at baseline, 2011)	0.40	0.45

Table 2

Individual patient level summaries

		SARI departments		Usual care departments	
		N/Mean	Percent/SD	N/Mean	Percent/SD
Patients included		32358	100	31915	100
Sex	Female	18374	56.8	16792	52.6
	Male	13984	43.2	15123	47.4
Age		54.0	SD=19.0	57.8	SD=18.5
Height		171.6	SD= 9.7	171.6	SD= 9.7
Weight		77.5	SD=18.3	77.0	SD=17.9
BMI		26.2	SD=5.5	26.1	SD=5.5
ASA class	ASA I	12293	38	9193	28.8
	ASA II	14624	45.2	14019	43.9
	ASA III	4730	14.6	7624	23.9
	ASA IV	436	1.3	918	2.9
	ASA V	23	0.1	51	0.2
	Unknown	252	0.8	110	0.3
Anticipated difficult intubation	Yes	1397	4.3	871	2.7
	No	30961	95.7	31044	97.3
Anticipated difficult mask ventilation	Yes	564	1.7	246	0.8
	No	31794	98.3	31669	99.2
Scheduled airway management plan	None/Unknown	405	1.3	54	0.2
	Spontanious breathing	306	0.9	217	0.7
	Mask ventilation	65	0.2	85	0.3
	Laryngeal Mask Airway	1415	4.4	1502	4.7
	Intubation with direct laryngoscopy	26825	82.9	28263	88.6
	Intubation with other methods	90	0.3	100	0.3
	Intubation with videolaryngoscope	2175	6.7	1467	4.6
	Tracheostomi in local anesthesia	4	0	0	0
	Fiber optic intubation	1073	3.3	227	0.7
Priority	Elective	23095	71.4	23328	73.1
	Emergency	9252	28.6	8557	26.8
Intubation	Maximum two intubation attempts - Only by direct laryngoscopy	29536	91.3	30060	94.2
	Maximum two intubation attempts in which other intubation equipment (e.g. video laryngo-scope) was used	2374	7.3	1417	4.4
	Three intubation attempts or more - Regardless of intubation method	423	1.3	424	1.3
	Intubation failed despite attempting	25	0.1	14	0
Mask ventilation	Not attempted	9299	28.7	9766	30.6
	Easy	22674	70.1	21810	68.3
	Difficult	351	1.1	307	1
	Impossible	24	0.1	9	0
Neuro muscular blocking agent	Not given	7305	22.6	3900	12.2
Anticipated difficult mask ventilation	Non-depolarizing	14491	44.8	16805	52.7
	Depolarizing	7150	22.1	7544	23.6
	Depolarizing and non-depolarizing	3403	10.5	3646	11.4
Stratum	High (≥ 2 % unanticipated difficult intubation at baseline, 2011)	17983	55.6	16592	52.0
	Low (< 2% unanticipated difficult intubation at baseline, 2011)	14375	44.4	15323	48.0

Table 3

Diagnostic accuracy for different populations and sensitivity analyses

The diagnostic accuracy of the anesthesiologists' prediction of difficult intubation

Population 1

- Patients that were attempted intubated but not pre-operatively scheduled for advanced intubation methods

SARI departments

		Difficult intubation		
		Yes	No	
Anticipated difficult intubation	Yes	82	415	497
	No	696	28016	28712
		778	28431	29209

Sensitivity	=	0.11	(0.09-0.13)
Specificity	=	0.99	(0.98-0.99)
Positive predictive value	=	0.16	(0.13-0.20)
Negative predictive value	=	0.98	(0.97-0.98)
Positive likelihood ratio	=	7.2	(5.8-9.1)
Negative likelihood ratio	=	0.9	(0.9-0.9)

Usual care departments

		Difficult intubation		
		Yes	No	
Anticipated difficult intubation	Yes	71	302	373
	No	723	29209	29932
		794	29511	30305

Sensitivity	=	0.09	(0.07-0.11)
Specificity	=	0.99	(0.99-0.99)
Positive predictive value	=	0.19	(0.15-0.23)
Negative predictive value	=	0.98	(0.97-0.98)
Positive likelihood ratio	=	8.7	(6.8-11.2)
Negative likelihood ratio	=	0.9	(0.9-0.9)

The diagnostic accuracy of the anesthesiologists' prediction of difficult intubation

Population 2

- Patients included as true positive if they were anticipated difficult to intubate, scheduled for and intubated by an advanced technique

SARI departments

Anticipated difficult intubation		Difficult intubation: Yes	Difficult intubation: No	
	Yes	852	413	1265
	No	694	27975	28669
		1546	28388	29934

Sensitivity	=	0.55	(0.53-0.58)
Specificity	=	0.99	(0.98-0.99)
Positive predictive value	=	0.67	(0.65-0.70)
Negative predictive value	=	0.98	(0.97-0.98)
Positive likelihood ratio	=	37.9	(33.1-42.1)
Negative likelihood ratio	=	0.5	(0.4-0.5)

Usual care departments

Anticipated difficult intubation		Difficult intubation: Yes	Difficult intubation: No	
	Yes	474	300	774
	No	723	29178	29901
		1197	29478	30675

Sensitivity	=	0.40	(0.37-0.42)
Specificity	=	0.99	(0.99-0.99)
Positive predictive value	=	0.61	(0.58-0.65)
Negative predictive value	=	0.98	(0.97-0.98)
Positive likelihood ratio	=	38.9	(34.1-44.4)
Negative likelihood ratio	=	0.6	(0.6-0.6)

The diagnostic accuracy of the anesthesiologists' prediction of difficult intubation

Sensitivity analysis 1

- All patients included and difficult intubation defined as an intubation score ≥ 3

SARI departments

Anticipated difficult intubation		Difficult intubation		
		Yes	No	
	Yes	73	1324	1397
	No	375	30586	30961
		448	31910	32358

Sensitivity	=	0.16	(0.13-0.20)
Specificity	=	0.96	(0.96-0.96)
Positive predictive value	=	0.05	(0.04-0.07)
Negative predictive value	=	0.99	(0.99-0.99)
Positive likelihood ratio	=	3.9	(3.2-4.9)
Negative likelihood ratio	=	0.9	(0.8-0.9)

Usual care departments

Anticipated difficult intubation		Difficult intubation		
		Yes	No	
	Yes	53	818	871
	No	385	30659	31044
		438	31477	31915

Sensitivity	=	0.12	(0.09-0.16)
Specificity	=	0.97	(0.97-0.98)
Positive predictive value	=	0.06	(0.05-0.08)
Negative predictive value	=	0.99	(0.99-0.99)
Positive likelihood ratio	=	4.7	(3.6-6.1)
Negative likelihood ratio	=	0.9	(0.9-0.9)

The diagnostic accuracy of the SARI versus the anesthesiologists' prediction of difficult intubation

Sensitivity analysis 2

- Per protocol analysis. The diagnostic accuracy of the SARI score compared with usual care departments. A SARI ≥4 represents anticipation of a difficult intubation in the SARI group.

SARI departments

		Difficult intubation		
		Yes	No	
SARI score ≥ 4	Yes	138	1139	1277
	No	496	21053	21549
		634	22192	22826

Sensitivity	=	0.22	(0.19-0.25)
Specificity	=	0.95	(0.95-0.95)
Positive predictive value	=	0.11	(0.09-0.13)
Negative predictive value	=	0.98	(0.97-0.98)
Positive likelihood ratio	=	4.2	(3.6-5.0)
Negative likelihood ratio	=	0.8	(0.8-0.9)

Usual care departments

		Difficult intubation		
		Yes	No	
Anticipated difficult intubation	Yes	71	302	373
	No	723	29209	29932
		794	29511	30305

Sensitivity	=	0.09	(0.07-0.11)
Specificity	=	0.99	(0.99-0.99)
Positive predictive value	=	0.19	(0.15-0.23)
Negative predictive value	=	0.98	(0.97-0.98)
Positive likelihood ratio	=	8.7	(6.8-11.2)
Negative likelihood ratio	=	0.9	(0.9-0.9)

Supplementary table

	GEE adjusted for cluster and stratum			GEE adjusted for cluster, stratum and potential confounders*			Unadjusted		
	Odds ratio (SARI/Usual care)	95% confidence intervals	P value	Odds ratio (SARI/Usual care)	95% confidence intervals	P value	Odds ratio (SARI/Usual care)	95% confidence intervals	P value
Population 1									
Unanticipated difficult intubation	1.03	0.77-1.38	0.84	1.17	0.81-1.67	0.40	1.00	0.90-1.11	0.98
Unanticipated easy intubation	1.26	0.68-2.34	0.47	1.43	1.06-1.92	0.02	1.43	1.23-1.66	<0.001
Population 2									
Unanticipated difficult intubation	1.02	0.77-1.35	0.89	NA	NA	NA	0.98	0.89-1.09	0.75
Unanticipated easy intubation	1.26	0.69-2.31	0.46	NA	NA	NA	1.42	1.22-1.65	<0.001
Sensitivity analysis 1									
Unanticipated difficult intubation	1.01	0.77-1.39	0.96	NA	NA	NA	0.96	0.83-1.11	0.58
Unanticipated easy intubation	1.19	0.72-1.97	0.50	NA	NA	NA	1.62	1.48-11.67	<0.001
Sensitivity analysis 2									
Unanticipated difficult intubation	0.95	0.70-1.28	0.72	NA	NA	NA	0.91	0.81-1.02	0.10

Population 1: Patients that were attempted intubated but not pre-operatively scheduled for advanced intubation methods

Population 2: Patients that were attempted intubated but not pre-operatively scheduled for advanced intubation methods plus patients anticipated difficult to intubate, scheduled for, and attempted intubated by an advanced method

Sensitivity analysis 1: All patients included and difficult intubation defined as an intubation score ≥ 3

Sensitivity analysis 2: The diagnostic accuracy of the SARI score compared with usual care departments

****Age; gender; ASA classification; emergency/elective procedure; body mass index; and use of neuromuscular blocking agents***