PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Evaluation of the McGrath MAC and Macintosh Laryngoscope for tracheal intubation in 2000 patients undergoing general anaesthesia: The randomised multicentre EMMA trial study protocol
AUTHORS	Kriege, Marc; Alflen, Christian; Tzanova, Irene; Schmidtmann, Irene; Piepho, Tim; Noppens, Ruediger

VERSION 1 - REVIEW

REVIEWER	Stephen Michael Kinsella Department of Anaesthesia St Michael's Hospital Bristol UK
REVIEW RETURNED	12-Oct-2016

GENERAL COMMENTS	This is a large enough study to achieve an answer to an important outcome – first-time tracheal intubation success rate – while also recording complications that might negate any benefit from the primary outcome if sufficiently frequent. Currently much advice / clinical practice is that videolaryngoscopy is used either after failure of tracheal intubation with the Macintosh direct laryngoscope, or in cases predicted to present a difficult intubation. This study is suggesting the use of videolaryngoscopy as a first line technique, in comparison to the current 'default' Macintosh laryngoscope. It should be emphasised that this study is being performed in selected cases who are low-risk for intubation difficulty - probably even in the title as well as the Introduction. The study is designed to recruit consecutive patients requiring tracheal intubation. It is acknowledged in the Methods that experience with procedures and devices takes a long time to develop; however the protocol does not specify the previous experience of the intubator with either device. As currently worded, there is the potential for bias should, for instance, a novice with the McGrath be presented with a patient who is predicted to present a difficult intubation (although lying just within the inclusion boundaries), leading to a more senior colleague performing the intubation instead. There should be more specific wording to address these aspects. Analysis Clarify those outcome measures that will be recorded by the observer, and those that are recorded by the intubator. The observer should be independent in this regard. Statistics

	There should be correction for multiple comparisons.

REVIEWER	TM COOK RUH bath UK My department of anaesthesia has received free or at cost equipment for research and evaluation from numerous airway companies. I have not made any profit of any sort from such donations. Neither I, not any members of my family have, to the best of our knowledge, any financial involvement in any airway companies. I have lectured at a Storz internal meeting, a competitor of Covidien, and attended a Covidien discussion group on capnography, both without being paid.
REVIEW RETURNED	18-Oct-2016

GENERAL COMMENTS	Evaluation of McGRATH MAC and Macintosh Laryngoscope for tracheal intubation in 2000 patients undergoing general anaesthesia: The randomised multicentre EMMA trial study protocol Kreige M et al. Review 18 October 2016 Comments The protocol does not state who the operators are or their training an experience with both devices. Inexperience (in general and specifically) and lack of training with the devices can bias the study significantly. This needs to be addressed. Are any post-op outcomes being measured? Sample size. Your predicted estimates of successful FPS need to be referenced against something that is a reliable clear source of that figure. There have been several published studies with the McGrath Mac so FPS estimates should be available from these papers or the authors – failing that from a pilot study. A poor power study may lead to as poor or even pointless study. The power calculation is also one sided; only considering that the McGrath will out-perform the Macintosh. That was not the case is a recent study. (Wallace CD, Foulds LT, McLeod GA, Younger RA, McGuire BE. A comparison of the ease of tracheal intubation using a McGrath MAC laryngoscope and a standard Macintosh laryngoscope. Anaesthesia. 2015; 70: 1281-5.) Can the authors please address this? We need more good quality evaluations of videolaryngoscopes. This study is large and well designed. Exclusion is balanced between the benefits of excluding known difficult patients (which would in theory expose the patients to risk due to multiple intubation attempts) and the down-sides of only studying easy patients (in which case it is likely that true benefits of the McGrath might be missed as it is designed to 'come into its own' when intubation is awkward/difficult).

	On balance I would think the researchers might include 'all those patients in whom intubation using a Macintosh laryngoscope blade is planned if they were not in the study' -– only exclude those patients in whom non-standard approaches to intubation were planned. This would capture more of the slightly difficult patients and thus make the results more generalizable and increase the likelihood of identifying a difference in performance between the two devices. It would replicate normal practice better. I do not think the study design would put these patients at risk that is higher than would have been the case if 'routine care' were employed, especially as the design is not cross-over. It is not clear whether the McGrath is to be used as a direct laryngoscope, as a videolaryngoscope or both. This should be clearly stated in the protocol. The Cormack and Lehane score is rather non-discriminatory as a measure of difficulty at laryngoscopy and I would recommend that a more discriminatory scale is used (e.g. as described in Cook TM. A new practical classification of laryngoscopy. Anaesthesia 2000; 55: 274-9). A fair amount of language subediting is needed. Lesser comments In the introduction the authors state a 1st pass success (FPS) rate with Macintosh laryngoscopy of 90%. Refs 11-13 are used to support the rate of FPS: Ref 12 is incomplete. It is an abstract and no paper seems to have followed. A FPS rate of 63% is not consistent with clinical practice. Ref 13 This reports a 79% FPS in an unselected population. Ref 14. The patient population was 'Patients were included if one or more of the following objective predictors of potentially difficult tracheal intubation were identified'. Despite this FPS was 84%. So I am uncertain on what basis the prediction of a 90% success rate in an unselected population is made. I note that 95% of patients had 1st or 2nd pass success is >90,000 intubations in the Danish Intubation Database (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; 110: 266-74) P7 bottom States 'The size of the endotracheal tube and the size of the blade are dependent on the standard operating procedure of the hospital' I think this should be standardised. E.g. size 4 used for males >1.85m or >90 kg. The danger is that the size 3 and 4 McGrath blades will not be used equally as the size 3 and 4 Macintosh blade, which could create further bias in an open label study. Similarly, size of tracheal tube impacts on ease of intubation so this should also be protocolised. P8 In an unblinded study it is possible that operator bias may lead them to try harder with one device and therefore succeed more despite unequal performance. How is this managed? For instance, does

	FPS have to be within 60 seconds which would be practical and pragmatically reasonable? I note on the Research registry it states intubation is limited to 120 s but this is not in the protocol. 60s seems a better cut-off. 'Time to view' is this time to 'any view of the larynx' or to 'best view of the larynx'. Please define. L28 After two failures does the operator 'change' or 'is advised to change' to other technique? This needs to be clearer. Secondary analyses – this is unclear. Collecting data on demographics (characteristics) in important but I'm not clear whether the authors are planning a sensitivity analysis by patient characteristics? P10 The description of allocation and sequence generation is not clear. There is only need to randomise to two groups. If block randomisation is used please say so and the size of the blocks. The protocol should state WHEN randomisation is performed. The operator should not know which device he/she will use until after anaesthetic induction. P11 Para 2 I don't see a need to perform statistics to compare randomly assigned groups' patient characteristics. It has crept in to mainstream but it is unnecessary if randomisation is done well. Para 4. This is the first mention of Likert scales? Measuring what? P12 Harms The protocol should describe how adverse events are reported to the ethical committee. Ethics Is the study also occurring in Canada? Table 1 I'm not sure this adds a lot as it includes numerous studies involving videolaryngoscopes other than the McGrath MAC.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

It should be emphasised that this study is being performed in selected cases who are low-risk for intubation difficulty - probably even in the title as well as the Introduction.
Thank you for pointing this out. We agree with the reviewer and have included this item in the "Introduction" section.

The study is designed to recruit consecutive patients requiring tracheal intubation. It is acknowledged in the Methods that experience with procedures and devices takes a long time to develop; however

the protocol does not specify the previous experience of the intubator with either device. As currently worded, there is the potential for bias should, for instance, a novice with the McGrath be presented with a patient who is predicted to present a difficult intubation (although lying just within the inclusion boundaries), leading to a more senior colleague performing the intubation instead. There should be more specific wording to address these aspects.
We thank the reviewer for this comment. We are aware of this possible bias and therefore will record the status (resident, specialist, consultant) as well as the previous experience of the provider with either DL or VL. We plan to evaluate first attempt success and the impact of experience with the device used. We now address this critical point in the “Study setting” section.

Analysis Clarify those outcome measures that will be recorded by the observer, and those that are recorded by the intubator. The observer should be independent in this regard.
Thank you for pointing this out. We wholeheartedly agree with the reviewer and have already taken precautions to address this issue. In all cases, an additional individual who is not involved in patient care (either a postgraduate student or a study nurse) is present during the induction of anaesthesia to records the study parameters. The description is now more accurate, and this point is addressed in the methods section.

Statistics There should be correction for multiple comparisons.
This is an important point! After consulting our statistician, we have included multiple comparisons of the primary outcome.

Reviewer 2

The protocol does not state who the operators are or their training an experience with both devices. Inexperience (in general and specifically) and lack of training with the devices can bias the study significantly. This needs to be addressed.
We wholeheartedly agree with the reviewer and now address this critical point in the manuscript. Please also see our comments to reviewer 1.

Are any post-op outcomes being measured?
We do plan to evaluate post-operative outcomes, such as a sore throat, and we do not have the consent to do so from our local ethics committee. However, acute complications (e.g., soft tissue trauma) are evaluated during anaesthesia induction. In such an extensive study, we intended to stay focused on the central question. In Germany, most patients are seen post-operatively by an anaesthesiologist, and common complications are documented in the patient's chart. Possibly, after ethical approval, we could do such an analysis retrospectively in the future.

Sample size. Your predicted estimates of successful FPS need to be referenced against something that is a reliable clear source of that figure. There have been several published studies with the McGrath Mac so FPS estimates should be available from these papers or the authors – failing that from a pilot study.
This is an interesting aspect. Studies comparing McGrath MAC with conventional direct laryngoscopy are currently rare. We evaluated two studies comparing the McGrath MAC and traditional laryngoscopy with an FPS of 91 – 92.8% (Purugganan J Cardiothorac Vasc Anesth. 2012; Kasuya Masui 2015; Alvis Minerva Anesth 2016). We appreciate that the methodological requirements (e.g., application in thoracic anaesthesia with double lumen tube, exclusively trainees with less experience and comparing McGrath MAC with KingVision) are very heterogeneous. In some countries, a consultant is always present during intubation, and in others, a resident is performing the procedure alone. Multiple factors could potentially influence sample size calculation for such a study. Many researchers in this field might have experienced difficulty in simply transferring study results to their

own OR. Therefore, adding the results of more studies might not have been helpful for increasing the precision based on the sample size in our specific setting.

A poor power study may lead to as poor or even pointless study. The power calculation is also one sided; only considering that the McGrath will out-perform the Macintosh. That was not the case is a recent study. (Wallace CD, Foulds LT, McLeod GA, Younger RA, McGuire BE. A comparison of the ease of tracheal intubation using a McGrath MAC laryngoscope and a standard Macintosh laryngoscope. Anaesthesia. 2015; 70: 1281-5.) Can the authors please address this?
We fully agree with the reviewer! We are aware of the study results of the cited manuscript. In contrast to this study, our protocol allows video laryngoscopic view of the glottis and video-guided intubation. Based on previous experiences with the device, we hypothesise a higher success rate in the first trial using the McGrath. We now include more detail about the sample size calculation in the methods section.

We need more good quality evaluations of videolaryngoscopes. This study is large and well designed. Exclusion is balanced between the benefits of excluding known difficult patients (which would in theory expose the patients to risk due to multiple intubation attempts) and the down-sides of only studying easy patients (in which case it is likely that true benefits of the McGrath might be missed as it is designed to 'come into its own' when intubation is awkward/difficult). On balance I would think the researchers might include 'all those patients in whom intubation using a Macintosh laryngoscope blade is planned if they were not in the study' -– only exclude those patients in whom non-standard approaches to intubation were planned. This would capture more of the slightly difficult patients and thus make the results more generalizable and increase the likelihood of identifying a difference in performance between the two devices. It would replicate normal practice better. I do not think the study design would put these patients at risk that is higher than would have been the case if 'routine care' were employed, especially as the design is not cross-over.
Thank you very much for this excellent comment! Several aspects led to the formulation of the inclusion criteria of this study. Our local ethics committee is very strict about inclusion criteria, which results in the maximum safety for study patients. We made several study proposals addressing expected difficult airway situations, and the committee refused to approve these proposals. Additionally, our institution is very strong in teaching awake fibreoptic intubations; currently approximately 7% of all intubations are performed using flexible intubation endoscopy. The general practice is that the threshold for an awake intubation is very low. Therefore, we currently only recruit patients who are not receiving fibreoptic intubation, which is very close to what the reviewer proposed.

It is not clear whether the McGrath is to be used as a direct laryngoscope, as a videolaryngoscope or both. This should be clearly stated in the protocol.
Thank you very much for noticing this. The provider is allowed to use direct or indirect vision according to his / her choice. This choice is well documented in the evaluation form. We have noticed that nearly every anaesthesiologist uses the McGrath solely as a video laryngoscope. Therefore, we expect a very low rate of direct visualisation using the McG. We have addressed this issue in the "secondary outcomes" section.

The Cormack and Lehane score is rather non-discriminatory as a measure of difficulty at laryngoscopy and I would recommend that a more discriminatory scale is used (e.g., as described in Cook TM. A new practical classification of laryngoscopy. Anaesthesia 2000; 55: 274-9).
We fully agree with the reviewer! Evaluation using the Cormack and Lehane score is a standard classification in our hospital and throughout Europe. C&L is accepted and used in multiple trials and manuscripts. We are well aware that C&L has multiple limitations. The POGO score is also an established and validated way to measure visualisation of the glottis. The "Intubation Difficulty Score"

was introduced some time ago and is a combination of subjective and objective criteria that potentially allows a qualitative and quantitative description of a challenging intubation. Consequently, we evaluate POGO and the Intubation Difficulty Score in this trial.

A fair amount of language subediting is needed.
We deeply apologise for the current form. We have used a scientific editing service in order to improve the language. We have included a certificate for the manuscript revision.

In the introduction the authors state a 1st pass success (FPS) rate with Macintosh laryngoscopy of 90%. Refs 11-13 are used to support the rate of FPS: Ref 12 is incomplete. It is an abstract and no paper seems to have followed. An FPS rate of 63% is not consistent with clinical practice. Ref 13 This reports a 79% FPS in an unselected population. Ref 14. The patient population was 'Patients were included if one or more of the following objective predictors of potentially difficult tracheal intubation were identified'. Despite this FPS was 84%. So I am uncertain on what basis the prediction of a 90% success rate in an unselected population is made. I note that 95% of patients had 1st or 2nd pass success is >90,000 intubations in the Danish Intubation Database (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 Anaesthesia Database. Anesthesiology. 2009; 110: 266-74)
We thank the reviewer for this critical comment. We have updated the references to address this issue. Based on our consultation with a statistician (Irene Schmidtmann, Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Centre of the Johannes Gutenberg-University, Mainz, Germany), sample size was calculated to show a difference of 5% between DL and VL.

P7 bottom States 'The size of the endotracheal tube and the size of the blade are dependent on the standard operating procedure of the hospital' I think this should be standardised. E.g. size 4 used for males >1.85m or >90 kg. The danger is that the size 3 and 4 McGrath blades will not be used equally as the size 3 and 4 Macintosh blade, which could create further bias in an open label study. Similarly, size of tracheal tube impacts on ease of intubation so this should also be protocolised.
Thank you for mentioning this important point. To our knowledge, no data exists in the literature recommending a detailed blade size correlated with the height of the patient. Our clinical standard recommends a #4 blade for very "tall" patients. The two centres have standardised tube sizes for use in patients: an ETT size 7.0 I.D. is currently used for females, and size 7.5 I.D for males. We now address this point in the "Methods" section.

P8 In an unblinded study it is possible that operator bias may lead them to try harder with one device and therefore succeed more despite unequal performance. How is this managed?
This is a very important point! There might be anaesthesiologists who prefer one device over the other and therefore try harder with their favourite device. Despite recruiting a large number of patients and including as many providers as possible, we do not have a method to rule out this potential bias.
For instance, does FPS have to be within 60 seconds which would be practical and pragmatically reasonable? I note on the Research registry it states intubation is limited to 120 s but this is not in the protocol. 60s seems a better cut-off.
We thank the reviewer for this comment. It is correct that we limit the time for intubation to 120 seconds. We now include this information in the manuscript. We respectfully disagree with the reviewer about the cut-off time for an intubation attempt. We are aware that many authors use this time as a cut-off for their, often very specific, setting. Since we evaluate a broad range of anaesthesiologists, including beginners, 60 seconds is too short to give less experienced practitioners enough time to visualise the glottis and place the endotracheal tube. We do not believe that we put the patient in any danger since every patient is well pre-oxygenated and we confirm sufficient ventilation before an intubation attempt.

‘Time to view’ is this time to ‘any view of the larynx’ or to ‘best view of the larynx’. Please define.
Thank you for this relevant comment! We evaluate the C&L class and POGO score at the moment when the “best view of the laryngeal structures” is obtained. In this context, we document whether further laryngeal manipulation (e.g., BURP) is necessary to achieve a C&L I / II. We are now more accurate in the “Methods” section.

L28 After two failures does the operator ‘change’ or ‘is advised to change’ to other technique? This needs to be clearer.
Thank you for noticing this! It is the policy of the department that after two unsuccessful intubation attempts, the technique has to be changed. The same applies for this study. Therefore, after two failures, the operator changes intubation technique. We are now more specific in the “Methods” section.

Secondary analyses – this is unclear. Collecting data on demographics (characteristics) in important but I’m not clear whether the authors are planning a sensitivity analysis by patient characteristics?
We do not plan to perform a sensitivity analysis of patient characteristics. We have corrected this point accordingly in the “Statistics” section.

P10 The description of allocation and sequence generation is not clear. There is only need to randomise to two groups. If block randomisation is used please say so and the size of the blocks. The protocol should state WHEN randomisation is performed. The operator should not know which device he/she will use until after anaesthetic induction.
The reviewer is correct. Please excuse that we did not clearly describe the randomisation process. A randomisation list was generated that randomly assigns individuals to the procedural group. We by randomising in groups of 200 (100 per group) for each study centre. We agree with the reviewer that it would be best if the operator is informed as late as possible about the study device before the intubation. Unfortunately, this is not possible due to organisational reasons. Each induction occurs in the "induction room", in which the anaesthesia nurse has the equipment prepared for the case. Before starting a case, the nurse and the anaesthesiologist both check the prepared equipment for presence and functionality (e.g., correct blade, stylet in tube). Therefore, the anaesthesiologist always knows what laryngoscope will be used before the patient arrives in the induction room.

P11 Para 2 I don’t see a need to perform statistics to compare randomly assigned groups’ patient characteristics. It has crept in to mainstream but it is unnecessary if randomisation is done well.
We fully agree with reviewer comment. Please also see the comment above (comment about secondary analysis).

Para 4. This is the first mention of Likert scales? Measuring what?
Thank you for noticing this! We use the “Likert scale” to assess the subjective degree of difficulty of the laryngoscopy for each individual (including glottic view, FPS, time to place the ETT). We address this now in more detail and have added the references to the “Methods” section.

P12 Harms The protocol should describe how adverse events are reported to the ethical committee.
This an important point. Reporting of severe adverse events (SAE) will be as per local Research Ethics Committee (REC) standard operating procedures. We address this now in the “Methods” section.

Ethics Is the study also occurring in Canada?
At this time point, the study is taking place at two hospitals in Germany. One of the authors recently

started an appointment in Canada and is currently looking into the possibility of starting another study centre. We are aware that further ethical approval and changes at clinicaltrials.gov are mandatory.

Table 1 I'm not sure this adds a lot as it includes numerous studies involving videolaryngoscopes other than the McGrath MAC.
We wholeheartedly agree with the reviewer. We now include all relevant studies focusing on McGrath MAC with other laryngoscopes in Table 1.

VERSION 2 – REVIEW

REVIEWER	TM Cook Royal United hospital, Bath UK My department of anaesthesia has received free or at cost equipment for research and evaluation from numerous airway companies. I have not made any profit of any sort from such donations. Neither I, not any members of my family have, to the best of our knowledge, any financial involvement in any airway companies. I have lectured at a Storz internal meeting, a competitor of Covidien, and attended a Covidien discussion group on capnography, both without being paid.
REVIEW RETURNED	02-Jan-2017

GENERAL COMMENTS	Evaluation of McGRATH MAC and Macintosh Laryngoscope for tracheal intubation in 2000 patients undergoing general anaesthesia: The randomised multicentre EMMA trial study protocol R1 Kreige M et al. Review 2nd January 2017 Comments There are some improvements but I still have significant concerns. The authors have definitely not addressed all my previous comments. I would like clarity as to whether the study has started or not (P13 L40 states it has). If it has it is difficult to change the protocol and as such somewhat perverse to publish it – especially as the required changes may not be made. The text flops from future to past tense. Throughout it should be in the future tense stating what participants and researchers WILL do rather than what they have done. The McG should be McGrath throughout. Sample size The study should be two sided – the McGrath may perform less well than direct laryngoscopy. Please address this directly. Who has funded the study? Who purchased the McGrath laryngoscopes?

	The discussion ads little to the paper. P9 Para 2 What training in use of the McGrath is provided – more specific details please. How many patient uses as a minimum before enrolling patients? P6 L40 I don't understand why the authors are actively choosing to only study patients predicted to be easy. This is the area where least harm happens to patients and where there is least likely to be benefit of videolaryngoscopy. Please justify. The exclusion of patients with predicted awkward or difficult laryngoscopy in whom direct laryngoscopy during general anaesthesia is still the planned technique would make the study of considerably more robust and the results of much greater clinical value. P6 L26-37 All these arguments also apply to the Glidescope, C-Mac and Venner – so why was the McGrath chosen? P9 L22 I have significant concerns about including beginners in this study. What is the minimum experience required with each device to be an intubator in the study? The major concern is that it switches from being an evaluation of the device/technique to being an evaluation of the performance/skills of the operator. This will likely dilute and confuse the results. P10 L23 I suggest the McGrath should always be used as a videolaryngoscope first. If not it may ONLY be used for direct laryngoscopy and in that case the main purpose of the study has not been explored (i.e. to compare VIDEOlaryngoscopy with direct laryngoscopy). P14 L28 When is the intubator performed of group allocation? This should occur after anaesthesia induction. Otherwise it is possible that bias will occur. Specifics P3 L5 The study only examines the McGrath Mac so cannot compare whether videolaryngoscopy performs better then direct laryngoscopy – only whether videolaryngoscopy with the McGrath Mac perform better than direct laryngoscopy – or less well. The study must be two sided. P4 L11 'allows' should be 'enables'

	L27 and P16 L10 Here it states a 95% 1st pass success with the McGrath and 90% for Macintosh – elsewhere (P16 L10) it is 90% vs 85%. Refs 14 and 15 are used to support an 85% 1st pass success rate with the Macintosh – both are from the emergency department. ED intubation is usually a rapid sequence induction and difficult laryngoscopy is up to 5-10x more frequent than in the elective anaesthesia setting. Please justify your 85% statistic and support it with relevant studies from the same population you are studying – i.e. elective patients without predictors of difficulty. Unless you can do this your power study will be invalid and the study may be pointless as the Macintosh 1st pass success rate may be much higher than you predict. L36 I have previously suggested that the Cook scale of other extended versions of the Cormack and Lehane score should be used to assess the view of the glottis. They are more discriminative – especially for the more difficult end of the spectrum of laryngoscopies. This has not been adopted. Instead the POGO score has been added – this adds almost no value at al as it only examines laryngoscopies where the glottis is visible – i.e. grade 1 and 2a. please justify or change as previously suggested. P5 The evidence is always developing and it might be useful to also mention the recent Cochrane review here. Videolaryngoscopy versus direct laryngoscopy for adult patients requiring tracheal intubation. Cochrane Database of Systematic Reviews 2016, Issue 11. Art. No.: CD011136. DOI: 10.1002/14651858.CD011136.pub2. Lewis SR, Butler AR, Parker J, Cook TM, Smith AF. P6 Table 1 includes 17 RCTs of videolaryngoscopy. As the recent Cochrane review included 64 can the authors clarify how and why these 17 were selected and if necessary alter or remove the table – it is not really necessary for the protocol. P9 L7 How is ‘high risk of aspiration’ defined? P9 L38-42 Preoxygenation should be standardised for the study as hypoxia is an outcome. P10 L18 As it states here that TOF count should be 0 at the point of laryngoscopy there should be no need to analyse this later as is described. (P12 L28) P10 L53-59 I see no point in separating the time from seeing the larynx and the time taken to pass the tube. All that matters is the overall time to intubate. That is the only clinically relevant time. Separating the times likely introduces measuring inaccuracy and errors.

	P11 L3 Is a bougie or stylet allowed to aid intubation? I would suggest for the second attempt it should be, to make the study more realistic and reduce risk to the patient from multiple laryngoscopies. This should be the same for both arms of the study. P12 L112 I don't think the Likert scale of ease is needed if the IDS score is already captured. P12 L28 See comment on P10 L18 P14 L34-53 This is probably not needed as it is routine. P15 L18 If randomisation is performed correctly there should be no need to perform statistical tests on the two groups' baseline characteristics. P16 L24-44 This is probably not needed as it is routine. P18 L35-40 I do not believe this statement is accurate. I also see no value and significant downside to complex timing during intubation.

VERSION 2 – AUTHOR RESPONSE

Comments to the reviewer

We wish to thank the referees for their valuable time and great efforts to review this manuscript! We believe that by addressing their comments we improved the manuscript significantly.
Comments from the reviewers are in *italics*.

Reviewer 2:

I *would like clarity as to whether the study has started or not (P13 L40 states it has).*

Patient inclusion began in November 2016. This now stated in the manuscript.

The text flops from future to past tense. Throughout it should be in the future tense stating what participants and researchers WILL do rather than what they have done.

The tenses have been adjusted throughout the manuscript.

The McG should be McGrath throughout.

We changed accordingly.

Sample size: The study should be two sided – the McGrath may perform less well than direct laryngoscopy. Please address this directly.

We are aware of the study results of the cited manuscript. In contrast to this study our protocol allows video laryngoscopic view of the glottis and video guided intubation. Based on previous experiences with the device, we hypothesize a higher success rate in the first trial using the McGrath. We now include sample size calculation in more detail in the methods section.

Who has funded the study?

This study is supported by the department of anaesthesia of the Medical Center of the Johannes Gutenberg University. No additional external funding is present. This is stated in the manuscript.

Who purchased the McGrath laryngoscopes?

The Medical Center of the Johannes Gutenberg-University purchased eight McGrath video laryngoscopes.

The discussion ads little to the paper.

We discuss the current data on video laryngoscopy and discuss possible limitations.

P9 Para 2 What training in use of the McGrath is provided – more specific details please. How many patient uses as a minimum before enrolling patients?

Every anaesthesiologist participating in this study has a minimal clinical experience of > 25 applications with the DL and > 5 uses of a video laryngoscope. We address this point in the manuscript.

P6 L40 I don't understand why the authors are actively choosing to only study patients predicted to be easy. This is the area where least harm happens to patients and where there is least likely to be benefit of videolaryngoscopy. Please justify. The exclusion of patients with predicted awkward or difficult laryngoscopy in whom direct laryngoscopy during general anaesthesia is still the planned technique would make the study of considerably more robust and the results of much greater clinical value.

Several aspects lead to the formulation of the inclusion criteria of this study. Our local ethics committee is very strict about inclusion criteria, which results in a maximum safety for study patients. We made several study proposals addressing expected difficult airway situations – the committee

refused all proposals because of patients' safety concerns. Additionally, our institution is very strong in teaching awake fibre optic intubations – currently approximately 7% of all intubations are performed using flexible intubation endoscopy. It is daily practice that the threshold for an awake intubation is very low. Therefore, we currently only recruit patients that are not receiving fibre optic intubation.

P6 L26-37 All these arguments also apply to the Glidescope, C-Mac and Venner – so why was the McGrath chosen?

The McGrath Mac video laryngoscope is already available at the institutions participating in this study. The McGrath uses disposable blade in different sizes and therefore can be quicker used between patients without re-processing a blade.

P9 L22 I have significant concerns about including beginners in this study. What is the minimum experience required with each device to be an intubator in the study? The major concern is that it switches from being an evaluation of the device/technique to being an evaluation of the performance/skills of the operator. This will likely dilute and confuse the results.

We do not agree with this comment. By including 2000 patients we already adjust for different levels of expertise. Many studies only examining the results of experts in the field can not be re-produced throughout the world because many physicians can not reach expert level.

P10 L23 I suggest the McGrath should always be used as a videolaryngoscope first. If not it may ONLY be used for direct laryngoscopy and in that case the main purpose of the study has not been explored (i.e. to compare VIDEOlaryngoscopy with direct laryngoscopy).

The McGrath is used for the initial attempt as a video laryngoscope.

P14 L28 When is the intubator performed of group allocation? This should occur after anaesthesia induction. Otherwise it is possible that bias will occur.

Unfortunately, we cannot meet this demand because of organizational reasons. The anaesthesia nurse has the equipment prepared for each case prior to the start of the case. Before starting a case the nurse and the anaesthesiologist both check the prepared equipment for presence and functionality (e.g. correct blade, stylet in tube). Therefore, the anaesthesiologist always knows what laryngoscope will be used before the patient arrives. For economic reasons the time for anesthesia can not be prolonged because of the study.

P3 L5 The study only examines the McGrath Mac so cannot compare whether videolaryngoscopy performs better then direct laryngoscopy – only whether videolaryngoscopy with the McGrath Mac perform better than direct laryngoscopy – or less well. The study must be two sided.

We use the McGrath because it`s a videolaryngoscope using Macintosh-based blade like other similar devices (e.g. C-MAC, Venner APA MAC). The benefit of this device is the single use blade, so we can use the instrument for more patients per day. With consultation and recommendation of our statistician (Dr. Irene Schmidtmann, Institute of Medical Biostatistics, Epidemiology and Informatics, University Medical Centre of the Johannes Gutenberg-University, Mainz, Germany), we prefer to use

one sided test. We hypothetize the superiority of video laryngoscopy with the McGrath, in this case one-sided testing is justifiable.

P4 L11 'allows' should be 'enables'

We corrected accordingly.

L27 and P16 L10 Here it states a 95% 1st pass success with the McGrath and 90% for Macintosh – elsewhere (P16 L10) it is 90% vs 85%. Refs 14 and 15 are used to support an 85% 1st pass success rate with the Macintosh – both are from the emergency department. ED intubation is usually a rapid sequence induction and difficult laryngoscopy is up to 5-10x more frequent than in the elective anaesthesia setting. Please justify your 85% statistic and support it with relevant studies from the same population you are studying – i.e. elective patients without predictors of difficulty. Unless you can do this your power study will be invalid and the study may be pointless as the Macintosh 1st pass success rate may be much higher than you predict.

We follow the suggestion of the reviewer and change the relevant references with a comparable study setting. Data from our department support an initial success rate of 85% using direct laryngoscopy.

L36 I have previously suggested that the Cook scale of other extended versions of the Cormack and Lehane score should be used to assess the view of the glottis. They are more discriminative – especially for the more difficult end of the spectrum of laryngoscopies. This has not been adopted. Instead the POGO score has been added – this adds almost no value at al as it only examines laryngoscopies where the glottis is visible – i.e. grade 1 and 2a. please justify or change as previously suggested.

Evaluation using the Cormack and Lehane score is a standard classification in our hospital and throughout Europe. C&L is accepted and used in multiple trials and manuscripts. We are well aware that C&L has multiple limitations. The POGO score is also an established and validated way to measure visualisation of the glottis. The "Intubation Difficulty Score" was introduced some time ago and is a combination of subjective and objective criteria that potentially allows a qualitative and quantitative description of a challenging intubation. Consequently, we evaluate POGO and the Intubation Difficulty Score in this trial.

P5 The evidence is always developing and it might be useful to also mention the recent Cochrane review here. Videolaryngoscopy versus direct laryngoscopy for adult patients requiring tracheal intubation. Cochrane Database of Systematic Reviews 2016, Issue 11. Lewis SR, Butler AR, Parker J, Cook TM, Smith AF.

Now we include this reference to the manuscript.

P6 Table 1 includes 17 RCTs of videolaryngoscopy. As the recent Cochrane review included 64 can the authors clarify how and why these 17 were selected and if necessary alter or remove the table – it is not really necessary for the protocol.

We did not include every available study on video laryngoscopy. Only studies examining the aspect DL vs. video laryngoscopy were included.

P9 L7 How is 'high risk of aspiration' defined?

We now define the risk of aspiration in the section "eligibility criteria".

P9 L38-42 Preoxygenation should be standardised for the study as hypoxia is an outcome.

P 9: The two methods for preoxygenation are recommended by our standard operating procedures. Both methods are standardised and established in the institutions participating in this study.

P10 L18 As it states here that TOF count should be 0 at the point of laryngoscopy there should be no need to analyse this later as is described.

We agree we the reviewer and delete this secondary endpoint.

(P12 L28) P10 L53-59 I see no point in separating the time from seeing the larynx and the time taken to pass the tube. All that matters is the overall time to intubate. That is the only clinically relevant time. Separating the times likely introduces measuring inaccuracy and errors.

The precise distinction of the overall time in three time sequence might be different between the devices and impart knowledge of the specific intubation process. We will be recruiting a large number of patients to see differences and trends in the separate intubation times.

P11 L3 Is a bougie or stylet allowed to aid intubation? I would suggest for the second attempt it should be, to make the study more realistic and reduce risk to the patient from multiple laryngoscopies. This should be the same for both arms of the study.

P 12 Secondary endpoints. A stylet is allowed to aid intubation. Bougies are not commonly used in Germany.

P12 L112 I don't think the Likert scale of ease is needed if the IDS score is already captured.

The IDS is a calculated Score, the Likert Scale give more subjectively characteristics of the device and the intubation process. We are recording both types to obtained more discriminable data.

P12 L28 See comment on P10 L18 P14 L34-53 This is probably not needed as it is routine.

We prefer to leave it as is.

P15 L18 If randomisation is performed correctly there should be no need to perform statistical tests on the two groups' baseline characteristics.

We do not plan sensitivity analysis of patient characteristics.

P18 L35-40 I do not believe this statement is accurate. I also see no value and significant downside to complex timing during intubation.

We choose to not change these measurements. Video laryngoscopy is often associated with difficult tube placement. By measuring time of the different steps of video laryngoscopy we should be able to further identify possible challenges with this technique.

VERSION 3 – REVIEW

REVIEWER	Nobuyasu Komasawa Osaka Medical College, Japan
REVIEW RETURNED	24-Apr-2017

GENERAL COMMENTS	This protocol is complete and adequate for evaluating large number of patients.

REVIEWER	Erkan Göksu Akdeniz University School of Medicine Department of Emergency Medicine Antalya/Turkey
REVIEW RETURNED	08-May-2017

GENERAL COMMENTS	I would like to thank to the editors of BMJ for giving me such an opportunity to review this manuscript. I also thank to the authors of this manuscript for their effort. This is a multicenter, open label, patient-blinded randomized controlled trial comparing first- pass success rate of direct laryngoscopy and McGrath MAC in the operation room. I have several suggestions to the authors of this paper. P4 line 2 before 'video laryngoscopy' they should add 'McGrath MAC' as this paper only compares a specific video laryngoscope with DL. P4line 6 I think the major limitation of this manuscript is different experience level of participants. Although different level of practitioners reflect real life situation, for such a study design I also

	suggest authors to compare the participants according to their experience levels. P8line 15 I think there is a typo error 'Another centre' should be removed P9 line12-15 Authors mentioned that 'participants have a clinical experience with at least 25 intubations using DL and at least five intubations using video laryngoscope', but in the 'introduction section' p5 line 6-7 ' to achieve a learning curve with a 90% probability of performing a successful intubation, a practitioner needs more than 57 attempts'. So I have concern if this experience level will have negative impact while comparing the two devices P10 line 31-32 in the literature 'time to intubation' is defined in various ways. In the manuscript authors preferred to use confirmation of ETCO2. This measurement method may limit the use of data for future analysis such as in meta-analysis. P11 line 6-7 During intubation with video laryngoscopy the view on the screen is two-dimensional and a stylet is often helpful while directing the endotracheal tube regardless of the blade shape. I would suggest authors to use stylet with all the intubations to get rid of possible effect of intubations without stylet with video laryngoscope. P12 line 3 How did the authors define the 'number of laryngoscopy attempts'? In one of my previous study I defined an attempt as an introduction of the laryngoscope into the mouth and its removal regardless of whether an ET tube was inserted. If the authors state the definition it is better. Page 12 line 22 I am not sure, the type of surgery is a necessary data to collect. If the authors think that may interfere with neck mobility, they are already collecting this data. Page 14 line 27 Although authors defined serious adverse events. They did not define adverse events. I suggest them to add.

VERSION 3 – AUTHOR RESPONSE

Reviewer 2:
P4 line 2 before ‘video laryngoscopy’ they should add ‘McGrath MAC’ as this paper only compares a specific video laryngoscope with DL.
We thank the reviewer for pointing this. We add the type of the video laryngoscope in the abstract section.

P4line 6 I think the major limitation of this manuscript is different experience level of participants. Although different level of practitioners reflect real life situation, for such a study design I also suggest authors to compare the participants according to their experience levels.
We wholeheartedly agree with the reviewer. Comparing the participants with their specific level of training is an important secondary endpoint. We now include this point in section “outcome measurement and analysis”.

P8line 15 I think there is a typo error ‘Another centre’ should be removed
We correct this accordingly.

P9 line12-15 Authors mentioned that ‘participants have a clinical experience with at least 25 intubations using DL and at least five intubations using video laryngoscope’, but in the ‘introduction section’ p5 line 6-7 ‘ to achieve a learning curve with a 90% probability of performing a successful intubation, a practitioner needs more than 57 attempts’. So I have concern if this experience level will have negative impact while comparing the two devices
We fully agree with the reviewer that this work has several biases, which we mention in the limitation section of the manuscript. However, the results reflect real life experience in two large teaching hospitals. Many studies are very well designed but are carried out by airway experts/airway enthusiasts with a high level of expertise. This might cause that the results of some studies cannot be replicated in daily anaesthesia practice by less experienced personnel. Consequently, we include physicians with a lower level of training (≥ 25 intubations using DL) to built up the daily situation in a tertiary hospital. We address this important point in section “Strengths and limitations of this study.”

P10 line 31-32 in the literature ‘time to intubation’ is defined in various ways. In the manuscript authors preferred to use confirmation of ETCO2. This measurement method may limit the use of data for future analysis such as in meta-analysis.
Thank you very much for this excellent comment! Several studies defined the time to intubation as the time until the first endtidal CO2 was observed on the anaesthesia monitor (Jones et al., Anesth Anal, 2008; Sun et al., Br J Anaesth, 2005) or until the cuff of the Endotracheal tube was blocked (Nouruzi-Sedeh et al. 2009; Serocki et al. 2010). We, therefore, measure multiple time points during/after intubation. We believe that “time to place” represents “time to intubation”- which has been used in other studies. We followed the reviewers suggestion and renamed the term to “time to intubation” throughout the manuscript.

P11 line 6-7 During intubation with video laryngoscopy the view on the screen is twodimensional and a stylet is often helpful while directing the endotracheal tube regardless of the blade shape. I would suggest authors to use stylet with all the intubations to get rid of possible effect of intubations without stylet with video laryngoscope.
We wholeheartedly agree with the reviewer. We will use a stylet for all intubations with the McGrath. Now we include the use of the stylet with the McGrath in the “Interventions” section.

P12 line 3 How did the authors define the ‘number of laryngoscopy attempts’? In one of my previous study I defined an attempt as an introduction of the laryngoscope into the mouth and is removal regardless of whether an ET tube was inserted. If the authors state the definition it is better.

Thank you for this helpful comment. We are fully agreed with the reviewer's suggestion and corrected this point in the "Interventions" section.

Page 12 line 22 I am not sure, the type of surgery is a necessary data to collect. If the authors think that may interfere with neck mobility, they are already collecting this data.
We thank the reviewer for this remark. We note this point in the section "Outcome measures".

Page 14 line 27 Although authors defined serious adverse events. They did not define adverse events. I suggest them to add
We thank the reviewer for pointing this. Now we describe adverse events in the "Harms" section.