GRAFT

GENERIC RATING OF ALLOGRAFT FUNCTION POST TRANSPLANT

FARID FOROUTAN, HBSc

A Thesis submitted to the School of Graduate Studies in Fulfillment of the Requirements for the Degree in Doctor of Philosophy

McMaster University

November 2019

DESCRIPTIVE NOTE

McMaster University DOCTOR OF PHILOSOPHY (2019), Hamilton, Ontario

(Department of Health Research Methods, Evidence, and Impact)

TITLE: GRAFT – Generic Rating of Allograft Function post-Transplant

AUTHOR: Farid Foroutan

SUPERVISOR: Gordon H. Guyatt

PAGES: 271

ABSTRACT

Background

Research on the optimal management of deceased organ donors poses unique challenges including the fact that one deceased donor may provide up to 8 organs for transplantation. Measuring the post-transplant function of these organs – good or bad – represents an attractive way of deciding whether treatment of deceased donors is working well, or not so well. Function, however, is organ-specific. Therefore, to conduct the most efficient and informative research on deceased donor management, we need an outcome measure that works well in all organs. The new outcome measure is called Generic Rating of Allograft Function post-Transplantation (GRAFT).

Methods

In this thesis, I highlight the methods for developing the cardiac-specific version of the GRAFT instrument. The same methods, however, have and will be applied to other organ-specific versions. The work comprised various study designs and developed novel research tools, all of which have advanced the development of the GRAFT instrument. At first, we developed a simple conceptualization for the instrument. Through regular consultation with research methodologists, biostatisticians and clinical experts, we refined the fundamental conceptualization and then refined the generic instrument, itself. One key concept is that GRAFT ratings should correlate with one-year graft function. To maximize its utility, I developed a heart--specific guide for applying GRAFT in future studies, and other organ-specific guides are underway. Specifically, we developed these guides by identifying the most robust predictors of one-year graft function through the conduct of organ-specific systematic reviews and meta-

analyses of prognostic factors. The evidence from these reviews, in consultation with a focus group of organ-specific transplant physicians, lead to refinements of our guides. We subsequently conducted a mixed-methods user testing to assess reliability and usability of the organ-specific guides. In appraising the evidence informing the guides, we developed GRADE guidance and a novel absolute risk calculator to assess our certainty in the body of evidence on prognostic factors informing our guides.

Results

We developed a 6-point generic rating instrument for classification of graft function to be applied post-transplant across all major solid organs. We designed GRAFT to be applied at the time of discharge, 1-month post-transplant, or at the time of death (whichever occurs first). We classify function as 1) normal, 2A) impaired but likely to gain normal function, 2B) impaired and unlikely to gain normal function, 3A) severely impaired but likely to gain some function, 3B) severely impaired and unlikely to gain some function, and 4) irreversible graft failure. Clinical expert collaborators for each organ type confirmed face validity of the GRAFT instrument.

For all organs, we identified a number of prognostic factors that can guide users in classifying organ function post-transplant. In consultation with clinical experts, we determined that the most important factor is graft function as measured by left ventricular ejection fraction (LVEF) or right atrial pressure (RAP). Due to limitations with the quality and quantity of the evidence, however, the heart transplant experts did not rely on the results of their organ group's systematic review. In turn, we conducted a retrospective cohort study to calculate the best estimate of association between LVEF, RAP, and overall mortality post heart transplant.

For the cardiac version of GRAFT, user testing demonstrated high reliability (Kappa of 0.87, 95% CI 0.62 – 1.00) and acceptable usability (system usability score of 75, inter-quartile range of 72.5 – 80).

In the process, we developed and published GRADE guidance for assessing certainty in the body of evidence addressing prognostic factors and devised a calculator to transform relative effect of each prognostic factor to absolute risks (http://hiru.mcmaster.ca/AbsoluteRiskCalculator/).

Conclusion

In this thesis, I advanced the development of an innovative generic instrument for the classification of graft function specifically for the purpose of application in clinical trials of deceased donor interventions. This work is ongoing, but very advanced for heart-specific components, for which I have ensured face validity, and demonstrated reliability and usability. The GRAFT instrument may better facilitate the conduct of future research to improve care of deceased organ donors with a view to improving quality and quantity of organs for transplantation.

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisors and mentors, Dr. Gordon Guyatt, Dr. Heather Ross, Dr. Maureen Meade, and Dr. Ana Carolina Alba. You have collectively and independently shaped me in to a better person. I am forever grateful to your teachings, kindness, patience, and guidance.

Gordon, thank you for the enormous depth of knowledge you've shared with me. You've truly inspired me to learn and grow, and I hope I can do the same for the next generation. Thank you for not only being my supervisor, but close and dear friend.

Heather, thank you for making this thesis possible. Without your support, encouragement, and funding, I would not have had any of the wonderful and unique experiences in the past four years. Thank you for showing me the world and pushing me to test my limits.

Caro, thank you for being my first contact with research and exemplifying what it means to be productive. I continue to learn and be inspired by you.

Maureen, thank you for trusting me with the development of GRAFT, exemplifying academia, sophistication, and philanthropy. Thank you for being supportive of the many tangents that came about in the last 4-years.

Special thanks to the funding sources: the Canadian Government from which I was awarded by the Vanier Canada Graduate Scholarships administered by the Canadian Institute of Health Research (CIHR) – Canada, Ted Rogers and the LaSorda family who funded my research fellowship in the Heart Failure and Transplant Program at the Toronto General Hospital.

I would like to thank all staff in the Department of Health Research Methods, Evidence, and Impact along with my classmates for enhancing experience in the last 4-years.

And thank you most of all, my loving and supportive family**.**

Table of Contents

Chapter 1: Introduction to this thesis

Optimization of medical care provided to deceased donors is an important strategy to maximize availability, longevity and performance of transplanted organs[1-3]. Currently, however, evidence to inform management strategies of deceased donors remains limited[3]. A particular challenge to designing studies to address optimal management is that one deceased donor may provide organs for up to 8 geographically separate recipients. Therefore, the impact of any intervention must consider all potential recipients from a single donor.

Currently, researchers with interest in studying management strategies in deceased donors may consider three outcome measures: 1) transplantation, 2) recipient/graft survival, 3) organ function post-transplant. Transplantation, as an outcome measure, presents sample size challenges, and does not provide insight regarding the function or longevity of the transplanted organs. Recipient or graft survival provides additional information beyond transplantation, but requires a long follow-up period that presents formidable feasibility challenges.

Organ function post-transplant represents an attractive intermediate outcome. How function is defined, however, is specific to each organ. For instance, heart, lung, and liver transplant, experts refer to early sub-optimal graft function as primary graft dysfunction. The specific definition of primary graft dysfunction for each group, and its relation to prognosis of transplant recipients is, however, completely different. The distinctive nature of each measure creates a challenge for researchers interested evaluating the impact of any therapy, applied to deceased donors, on all organs procured. If we develop a single measure of function that could be applied to each organ, it would allow us to address the overall impact of an intervention on all organs, and enhance the efficiency of studies of interventions in deceased organ donors.

To address the need for a single universal measure of function, we set forth to develop such a generic outcome measurement (**chapter 7**). We strived to make the instrument generic by linking all organ groups through prognosis. Therefore, the instrument we developed, and present in chapter 7, classifies patients based on expected graft longevity within the year post-transplant.

In the process of developing our instrument for heart transplantation, we explored the validity of currently available classification systems such as the International Society of Heart and Lung Transplantation's Primary Graft Dysfunction instrument (**chapter 1**)[4]. Through understanding and applying their instrument, we became aware of strengths and limitations that further guided and informed our development of GRAFT.

In pursuit of concomitantly developing GRAFT for other organs, we conducted systematic reviews and meta-analyses of prognostic factors to identify all donor, transplant, and recipient factors that influence graft longevity in the first-year post-transplant (**chapters 2, 3, and 4**)[5 6]. These factors in turn informed the thresholds used for classification of patients in each category of GRAFT.

As we conducted these reviews, we required guidance for determining our certainty in the impact of each prognostic factor on graft loss. Our level of certainty reflects on the trustworthiness of the direction and magnitude of associations calculated by the reviews. To this end, with the approval and support of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) guidance group we developed guidance for assessing certainty in the effect of prognostic factors (**chapter 5**).

As we developed our GRADE guidance for certainty, two broad uses of prognostic factors became clear. The first is in relation to study planning and analysis: stratification of

randomization, adjusted analysis, and developing a prognostic model. The second is in relation to clinical decision-making which implies that clinicians and patients need risk estimates defined by the prognostic factor to inform their decisions. Throughout chapters 2-5, it became evident that although relative effects are useful for understanding prognostic factors across a wide range of patients, when helping them in planning their care/future absolute risks become necessary. To this end, **chapter 6** presents an approach for calculation of absolute risks in those with and without prognostic factors[7].

References

1. Bos EM, Leuvenink HG, van Goor H, et al. Kidney grafts from brain dead donors: Inferior quality or opportunity for improvement? *Kidney Int* 2007;72(7):797-805. doi: 10.1038/sj.ki.5002400 [published Online First: 2007/07/27]
2. Floerchinger B, Oberhuber R, Tullius SG. Effects of brain death on organ quality and transplant outcome. *Transplant Rev (Orlando)* 2012;26(2):54-9. doi: 10.1016/j.trre.2011.10.001 [published Online First: 2012/03/31]
3. Greer DM, Valenza F, Citerio G. Improving donor management and transplantation success: more research is needed. *Intensive Care Med* 2015;41(3):537-40. doi: 10.1007/s00134-015-3661-0 [published Online First: 2015/01/27]
4. Foroutan F, Alba AC, Stein M, et al. Validation of the International Society for Heart and Lung Transplantation primary graft dysfunction instrument in heart transplantation. *J Heart Lung Transplant* 2018 doi: 10.1016/j.healun.2018.12.007 [published Online First: 2019/01/16]
5. Foroutan F, Friesen EL, Clark KE, et al. Risk Factors for 1-Year Graft Loss After Kidney Transplantation: Systematic Review and Meta-Analysis. *Clin J Am Soc Nephrol* 2019 doi: 10.2215/CJN.05560519 [published Online First: 2019/09/22]
6. Foroutan F, Guyatt G, Friesen E, et al. Predictors of 1-year mortality in adult lung transplant recipients: a systematic review and meta-analysis. *Syst Rev* 2019;8(1):131. doi: 10.1186/s13643-019-1049-x [published Online First: 2019/06/05]
7. Foroutan F, Iorio A, Thabane L, et al. Calculation of absolute risk for important outcomes in patients with and without a prognostic factor of interest. *Journal of clinical epidemiology* 2019;117:46-51. doi: 10.1016/j.jclinepi.2019.08.012 [published Online First: 2019/10/08]

Chapter 2: Validation of The International Society of Heart and Lung Transplantation (ISHLT) Primary Graft Dysfunction instrument in heart transplantation

STATUS: Published – The Journal of Heart and Lung Transplantation

https://doi.org/10.1016/j.healun.2018.12.007

Farid Foroutan[1,3], Ana Carolina Alba[1,3], Madeleine Stein[1], John Krakovsky[2], Kevin Gar Wai Chien[2], Sharon Chih[2], Gordon Guyatt[3], Heather Ross[1]

1- Ted Rogers Centre for Heart Research, Multi-Organ Transplant Program, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

2- Heart failure – transplantation, Division of Cardiology, University of Ottawa Heart Institute, Ottawa, Ontario, Canada

3- Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

ABSTRACT

Background

In 2014, the ISHLT developed a classification instrument for left ventricular (LV) and isolated right ventricular (RV) primary graft dysfunction post heart transplant. The instrument classifies LV-PGD as mild, moderate, or severe. This study evaluates the predictive validity of this instrument.

Methods

We conducted a cohort study of 412 consecutive patients transplanted between 2004 to 2015 at Toronto General Hospital and Ottawa Heart Institute (Canada). We classified LV-PGD as mild, moderate or severe using the ISHLT instrument. To assess predictive validity, we evaluated the association between LV-PGD severity and 1-year post-transplant mortality using a Cox-regression model adjusted for recipient age.

Results

The cohort was predominantly male (71%), mean age 50 ± 13, mean donor age 38 ± 14, with 25% female donors. Mean ischemic time was 3.7 ± 1.1 hours. Mild LV-PGD occurred in 3.6% of patients, moderate in 9.5%, and severe in 3.9%. All levels of LV-PGD were associated with increased 1-year mortality, with a gradient in association between mild, moderate and severe. We only observed a statistically significant association for moderate and severe form of LV-PGD (mild: HR 2.4, 95%CI 0.6–10.2; moderate: HR 7.0, 95%CI 3.4–14.6; severe: HR 15.9, 95%CI 7.2–35.0).

Conclusion

The ISHLT LV-PGD classification convincingly identifies a substantial increase in risk of death at one year, and an increased gradient of risk, in those with moderate or severe LV-PGD.

INTRODUCTION

The 2017 International Society of Heart and Lung Transplantation's (ISHLT) registry report described outcomes on 80,000 heart transplants performed from 1994 to 2015. The risk of 1-year mortality in this cohort was 16.5%. Of these, 5,400 (41%) deaths occurred within the first-year post transplant due to, or related to, graft dysfunction[1].

Early graft dysfunction without a clear precipitating cause, termed primary graft dysfunction (PGD), is infrequent but associated with significant morbidity and mortality. Literature published prior to 2014 utilized various definitions for PGD (ranging from the need for mechanical circulatory support within the first 30 days to use of low dose inotropes) and reported variable incidence rates (2.8% to 23%,)[2]. Variability in the definition of PGD created difficulty in comparing the risk of mortality over time and across centers, limiting potential for development of management strategies for such patients.

To address these challenges, in 2014 the ISHLT developed a consensus-based definition and classification instrument for PGD [3]. The society now defines PGD as graft dysfunction not due to rejection, volume overload, or pulmonary hypertension, that occurs within the first 24 hours post-transplantation. The consensus panel classified PGD as left ventricular (LV), bi-ventricular (BV)-PGD or isolated right ventricular (RV)-PGD. For LV and BV-PGD, the instrument further subclassifies dysfunction as mild, moderate, or severe (distinguished by the level of support required). Since the ISHLT publication, four studies that used the consensus definition observed 30-day mortality incidence of 30 to 51% in patients with severe PGD.[4].

For an instrument such as the ISHLT PGD classification, association with subsequent outcomes represents a potentially powerful demonstration of validity. To date, no study has formally assessed the performance of the severity classification of the PGD instrument when applied

within the first 24 hours post-transplant in relation to 1-year mortality. Therefore, we evaluated the association between the ISHLT classification of graft dysfunction and mortality during the first post-transplant year in heart transplant recipients.

METHODS

Population

We collected a cohort of 412 consecutive adult heart transplant recipients (January 1, 2004 to January 1, 2015), followed at Toronto General Hospital or Ottawa Heart Institute. We excluded patients under the age of 18, patients undergoing re-transplantation, and recipients of more than one organ.

Recipient Data

We abstracted age, sex, co-morbidities, and heart failure etiology, pre-transplant laboratory values, and pre-transplant support of transplant recipients, as well as inotropic support and mechanical circulatory support (intra-aortic balloon pump, extracorporeal membrane oxygenation, ventricular assist devices) post-transplant.

Donor Data

We abstracted donor age, sex, total ischemic time (clamp of aorta in donor, release of clamp in recipient), inotropic support (only available post 2013), echocardiographic assessment, cause of brain death, troponin (only available post 2013), and angiography (only available post 2013).

Primary Graft Dysfunction

We identified PGD using data from the first 24 hours post-transplant. We excluded other potential causes for graft dysfunction on the basis of the clinician team's consensus assessment and the results of specific test results (including, for example acute cellular rejection by endomyocardial biopsy). We recorded diastolic pulmonary artery pressure in the post-operative

period as an alternative when pulmonary capillary wedge pressure measures were unavailable[5]. We classified patients according to the ISHLT PGD instrument published by Kobashigawa et al. (Table 1). As per the instrument definition, we classified individuals experiencing both LV and RV dysfunction as LV-PGD. One individual blinded to the outcomes of recipients, considered all hemodynamic measures and supports recorded in medical records of patients within the first 24 hours. We classified patients with PGD at the first instance it was observed based on criteria proposed by ISHLT. We extracted all data necessary for classification of PGD from intensive care unit flow sheets and charts and echocardiographic reports.

Outcome

The primary outcome of our study was all-cause 1-year mortality. All patients were followed for 1-year without any loss to follow-up.

Statistical Analysis

We summarized continuous variables as mean and standard deviation (for normally distributed data) or median and inter-quartile range (IQR) (for skewed data) and categorical data using absolute counts and proportions. For comparison of continuous data across different PGD severity groups, we utilized one-way ANOVA. In cases of skewed continuous data, we utilized the non-parametric Dunn's test. To compare continuous data between patients with RV-PGD and no PGD, we utilized the two-way t-test. We used the chi-square test (or Fisher's exact test if number of patients in was less than five) to compare categorical data across no PGD, LV-PGD severity groups, and RV-PGD.

We depicted the observed survival using Kaplan Meier analysis, using the log-rank test to assess if chance could explain differences in observed survival across PGD severity classes. For prognostication and generalizability to future transplant recipients, we modelled the association

between severity levels of LV-PGD and all-cause 1-year mortality using Cox proportional hazard regression analysis. Each severity level of PGD was entered as a dummy variable with no-PGD as the reference group. Because of the small number of events, to avoid overfitting we restricted adjustment of the model to recipient age. We could not conduct similar analysis for RV-PGD due to low number of deaths. We tested the proportional hazard assumption graphically using the log-log plot of survival, and statistically using Schoenfeld residuals. All statistical analyses were conducted with STATA 15.1. A p-value <0.05 suggested statistical significance.

The research ethics boards of both University Health Network and Ottawa Heart Institute approved the study.

RESULTS

Table 2 presents the characteristics of the 412 eligible heart transplant recipients of whom 249 received their transplant at the Toronto General Hospital and 163 at the Ottawa Heart Institute. The majority of recipients were male (71%) with mean age of 50 $\pm$ 13. Bridging to transplant with left ventricular assist device occurred in 127 (32%) patients. Of the donors 67% were male with an average donor age of 38 $\pm$ 14 years; 25% of transplants were donor/recipient sex-mismatched. The mean ischemic time was 3.7 $\pm$ 1.1 hours. Table 3 summarizes the hemodynamic findings used for classification of PGD within the first 24 hours.

LV-PGD

The ISHLT system classified 15 (3.6%) patients as mild, 39 (9.5%) as moderate, and 16 (3.9%) as severe PGD (figure 1). Patients with severe PGD had longer ischemic time (4.4 $\pm$ 1.3 hours vs. 3.5 $\pm$ 1.1 in patients without PGD, p-value <0.001) and higher proportion of patients bridged to transplant with LVAD (67% vs. 31% in patients without PGD, p-value 0.03). The recipient, donor,

and transplant characteristics proved comparable across different LV-PGD severity levels (table 2). Of the 44 deaths in the first-year post transplant, retrospective chart review described cause of death as graft failure for 35%, multi-organ failure 28%, stroke 11%, sepsis 11%, infection 7%, and pulmonary bleed 2%. Almost all deaths occurred within the first 30-days post-transplant: 30-day survival of 97.6% (95% CI 95.2 – 98.8%) in those without LV-PGD, 86.7% (95% CI 56.4 – 96.5%) with mild LV-PGD, 71.8% (95% CI 54.9 – 83.3%) with moderate LV-PGD, and 47.4% (95% CI 21.8 – 69.4%) with severe LV-PGD (figure 2, log-rank p-value <0.001).

In a cox-regression model adjusted for recipient age, only moderate and severe levels of LV-PGD were significantly associated with 1-year mortality (table 4). We observed a graded increase in the strength of association between mild, moderate, and severe LV-PGD with wide confidence intervals. We excluded an association by chance (p-value <0.05) for moderate and severe LV-PGD. Our cox regression model met the proportional hazard assumption. In the baseline group of no LV-PGD, there was an absolute mortality risk of 6%. This risk increased to 16% (95% CI 3% – 52%) for mild, 40% (95% CI 22% – 65%) for moderate, and 68% (95% CI 41 – 92%) for severe PGD. In a separate cox-regression model (data not shown), we adjusted for the transplant center to ensure center-specific differences in treatment are not possibly influencing our primary outcome of 1-year mortality. Adjustment for this made no significant difference in the association between each LV-PGD severity and 1-year mortality.

RV-PGD

Isolated RV-PGD defined by hemodynamic criteria (N=11) or RVAD (N=1) occurred in 12 (3.5%) recipients. Patients with RV-PGD were predominantly male (58.3%) with a mean age of 52 $\pm$ 16 (table 1). Of the donors 50% were men; donor mean age was 35 $\pm$ 16. Patients with RV-PGD

were more frequently female (58%) compared to no PGD patients (27%). Two patients with RV-PGD died during a mean follow-up 11.6 $\pm$ 2.4 months. The observed 30-day survival in patients with RV-PGD was 92% (95% CI 54 – 99%) (p = 0.16 in comparison to patients without PGD).

DISCUSSION

Principle findings

Our findings demonstrate a strong association between the ISHLT PGD classification of LV dysfunction and 1-year mortality, particularly for those recipients with moderate or severe dysfunction (Figure 1), thus providing strong support for the validity of the LV component of the ISHLT system. Most deaths occurred in the first 30 days, with a gradient in deaths between the four categories (Figure 1). Isolated RV dysfunction was not associated with mortality, with very few deaths – thus, our results do not inform the validity of this aspect of the ISHLT system.

Relation to prior studies

Previous studies addressing the validity of the ISHLT PGD classification system modified the definition of PGD and its severity levels or aggregated the different severity levels[6-9]. Our study is the first to validate the severity classifications of PGD without modifying the originally published criteria. The age adjusted predicted risks of 1-year mortality in our study are similar to the unadjusted observed 1-year mortality reported by Dronavalli et al. (15%, 41%, and 67% in mild, moderate, and severe respectively)[9]. These authors reported validity of the ISHLT PGD instrument based on classification from the first 72 hours as opposed to the first 24 hours post-transplant lowering confidence in applicability and performance of the instrument when assessments are made within the first 24 hours. Dronavalli et al. observed no difference in the

survival of mild PGD patients, as compared to those without PGD. In our cohort, we observed a 2-fold increase in risk of mortality, though not statistically significant, in these patients.

In our study, we observed a statistically significant association between increasing recipient age and decreasing risk of 1-year mortality. Two previous studies also observed a protective association between increasing recipient age and mortality[10,11]. One possible explanation for the protective association may be due to the confounding effect of congenital heart disease as these patients are significantly younger compared to all other heart failure aetiologies, and at highest risk of 1-year mortality[1]. In addition, older heart transplant candidates are less likely to have comorbidities due to the careful candidacy selection process which could minimize their post-operative risk. The increased prevalence of comorbidities necessitates careful selection of candidates, relative to the higher risk tolerance for younger recipients[12].

Implications

The ISHLT classification for moderate and severe LV-PGD has important prognostic information for the first-year post-transplant. If a prediction model with adequate discrimination and calibration that identifies patients at high risk of developing moderate or severe cases of PGD were available, it could be useful for recipient selection and organ allocation.

To date, the most utilized risk prediction tool for early graft dysfunction is the RADIAL risk score. Authors of the RADIAL score defined early graft dysfunction by the presence of one of the following criteria: (1 - significant impairment of systolic function affecting both left, right or both ventricles; 2- severe hemodynamic compromise lasting over 1 hour: systolic BP <90 mmHg and/or CI <2.2 L/min/m^2 requiring two or more IV inotropes/pressor drugs, or MCS support, 3 – occurring within the firth 24 hours; 4 – absence of any other obvious cause). The RADIAL score's classification of early graft dysfunction does not categorize patient based on severity. Our study

shows important difference across severity levels of PGD. Therefore, there is need for future studies to develop novel risk models to predict not only PGD but also its severity as defined by the ISHLT instrument.

Strengths & Limitations

We designed our study to assess the predictive validity of the PGD severity classification, as originally proposed by the ISHLT, ensuring that our classification of PGD severity was blinded to recipient's outcomes. We enrolled a consecutive sample of patients, ensuring representativeness of the population, and achieved 100% follow-up at one year. we assessed predictive validity with use of Cox-proportional hazard model, which met the proportional hazards assumption and provided hazard ratios for mild, moderate, and severe LV-PGD.

Limited information required one minor modification to the instrument. PCWP was measured in only 65 patients from the Ottawa Heart Institute and 3 from Toronto General Hospital. In order to meet the PCWP criteria for classification of PGD, we therefore relied on diastolic PA pressure. Although diastolic PA is a good surrogate for PCWP in healthy individuals with normal pulmonary vasculature[5], its validity in the HT population remains unknown. Given that irreversible pulmonary hypertension is a relative contraindication to heart transplantation[12], diastolic PA may serve as an adequate surrogate in this population as well. Our results apply directly to all centers that do not use routine measurement of PCWP in the post-operative setting, and indirectly to those that do.

Only one individual, on the basis of chart review, made the decision that there was no other cause of cardiac dysfunction, and classified the PGD severity of patients using hemodynamic data from the first 24 hours post-transplant. Ideally, in order to assess replicability, these judgments should have been made by two independent individuals.

Only 70 patients met classification criteria for LV-PGD. Hence, our study lacked power to provide precise 95% confidence intervals when modelling the impact of each severity level on 1-year mortality. Although our model validates the use of the ISHLT instrument for classifying the severity of PGD, the observed imprecision decreases our confidence in the impact of each PGD severity on 1-year mortality[13].

The ISHLT PGD instrument is a valid tool for classifying the severity of graft dysfunction observed in the post-operative period. The instrument, however, utilizes both functional thresholds and medical therapy (in the form of inotropic agents) for classification of patients. The threshold for timing, type, and titer of inotropic support may vary considerably across physicians and programs. Such variability in practice may influence the associations between LV-PGD severity and mortality. Therefore, there is a potential limitation in applying our results across centers with different approaches to use of inotropes in patients with compromised cardiac function.

In our study, we did not capture information on the proportion of patients transplanted from ECMO, IABP, or inotropic support. The lack of information limits attempts to compare and contrast our cohort to other centers, to which our results may be applied to.

CONCLUSION

The ISHLT PGD instrument defines and classifies the severity of a well-recognized post-transplant occurrence. Prior to the publication of ISHLT's PGD instrument, variation in the classification of PGD created difficulty in understanding the possible underlying causes. With a valid consensus instrument, we can better study the risk factors, histology, and biological manifestation of PGD (all of which help in definitively understanding the underlying etiologies).

Our results provided support for the validity of the ISHLT's LV-PGD system and thus its use in future studies of primary graft dysfunction.

ACKNOWLEDGEMENTS

The authors acknowledge the Trillium Gift of Life Network for providing data on donors

DISCLOSURES

None of the authors have any conflicts to disclose.

REFERENCES

1. Lund LH, Khush KK, Cherikh WS, et al. The Registry of the International Society for Heart and Lung Transplantation: Thirty-fourth Adult Heart Transplantation Report-2017; Focus Theme: Allograft ischemic time. J Heart Lung Transplant 2017;36:1037-46.
2. Segovia J, Cosio MD, Barcelo JM, et al. RADIAL: a novel primary graft failure risk score in heart transplantation. J Heart Lung Transplant 2011;30:644-51.
3. Kobashigawa J, Zuckermann A, Macdonald P, et al. Report from a consensus conference on primary graft dysfunction after cardiac transplantation. J Heart Lung Transplant 2014;33:327-40.
4. Foroutan F, Ross HJ. Primary Graft Dysfunction: The Devil is in the details. Transplantation 2018.
5. Wright SP, Esfandiari S, Gray T, et al. The pulmonary artery wedge pressure response to sustained exercise is time-variant in healthy adults. Heart 2016;102:438-43.
6. Lima EB, Cunha CR, Barzilai VS, et al. Experience of ECMO in primary graft dysfunction after orthotopic heart transplantation. Arq Bras Cardiol 2015;105:285-91.
7. Sabatino M, Vitale G, Manfredini V, et al. Clinical relevance of the International Society for Heart and Lung Transplantation consensus classification of primary graft dysfunction after heart transplantation: Epidemiology, risk factors, and outcomes. J Heart Lung Transplant 2017;36:1217-25.
8. Squiers JJ, Saracino G, Chamogeorgakis T, et al. Application of the International Society for Heart and Lung Transplantation (ISHLT) criteria for primary graft dysfunction after cardiac transplantation: outcomes from a high-volume centredagger. Eur J Cardiothorac Surg 2017;51:263-70.
9. Dronavalli VB, Rogers CA, Banner NR. Primary Cardiac Allograft Dysfunction-Validation of a Clinical Definition. Transplantation 2015;99:1919-25.
10. Custódio IL, Lima FET, Lopes MVdO, et al. Results of medium-term survival in patients undergoing cardiac transplantation: institutional experience. Revista brasileira de cirurgia cardiovascular : orgao oficial da Sociedade Brasileira de Cirurgia Cardiovascular 2013;28:470-6.
11. Vakil K, Taimeh Z, Sharma A, et al. Incidence, predictors, and temporal trends of sudden cardiac death after heart transplantation. Heart Rhythm 2014;11:1684-90.
12. Mehra MR, Canter CE, Hannan MM, et al. The 2016 International Society for Heart Lung Transplantation listing criteria for heart transplantation: A 10-year update. The Journal of Heart and Lung Transplantation 2016;35:1-23.
13. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol 2011;64:1283-93.

TABLE 1 - Primary Graft Dysfunction classification

Ventricle	Severity	Criteria
LV or BiV PGD	Mild	LVEF ≤ 40%, or RAP >15 mmHg & PCWP >20 mmHg & CI <2.0 L/min2 (>1 hour) & requiring low dose inotropes
	Moderate	i) LVEF ≤ 40%, or RAP >15 mmHg & PCWP >20 mmHg & CI <2.0 L/min2 & MAP <70 mmHg (>1 hour)
		II) High dose inotropes, or newly place IABP
	Severe	Dependence on LV or Bi-Ventricular MCS (Excluding IABP)
RV-PGD	-	i) RAP >15 mmHg & PCWP <15 mmHg & CI <2.0 L/min2
		ii) TPG <15 mmHg, and/or PA systolic >50 mmHg
		iii) Need for RVAD

LV – Left Ventricular; BiV – Bi-Ventricular; PGD – Primary Graft Dysfunction; RV – Right Ventricular; LVEF – Left ventricular ejection fraction; PCWP – pulmonary capillary wedge pressure; CI – Cardiac Index; IABP – Intra-aortic balloon pump; MAP – mean arterial pressure; MCS – Mechanical circulatory support; RAP – Right atrial pressure; TPG – Transpulmonary gradient; PA – pulmonary artery; RVAD – Right ventricular assist device.
Inotrope score = dopamine (X1) + dobutamine (X1) + amrinone (X1) + milrinone (X15) + epinephrine (X100) + norepinephrine (X100) with each drug dosed in µg/kg/min

TABLE 2 - Baseline recipient, donor, and transplant characteristics

	All (n = 412)	None (n = 330)	Mild (n = 15)	Moderate (n = 39)	Severe (n = 16)	p-value	RV-PGD (n = 12)	p-value
Recipient Characteristics								
Age	50 ± 13	50 ± 13	49 ± 13	53 ± 11	49 ± 15	0.439	51 ± 12	0.667
Male sex	294 (71%)	240 (73%)	12 (80%)	27 (69%)	10 (62%)	0.706	5 (42%)	0.043
BMI	25 ± 5	25 ± 5	25 ± 5	26 ± 4	26 ± 5	0.638	24 ± 5	0.362
Pre-transplant Diabetes	81 (21%)	64 (20%)	6 (43%)	9 (25%)	2 (12%)	0.173	2 (17%)	1.000
Ischemic HF Etiology	115 (28%)	93 (28%)	2 (13%)	9 (23%)	7 (44%)	0.280	4 (33%)	0.747
Previous sternotomy	202 (50%)	162 (50%)	7 (47%)	20 (56%)	11 (73%)	0.326	2 (17%)	0.036
LVAD	127 (32%)	103 (31%)	3 (20%)	11 (30%)	10 (67%)	0.032	3 (25%)	0.762
ECMO	7 (2%)	6 (2%)	0 (0%)	0 (%)	1 (6%)	0.337	0 (0%)	1.000
IABP	3 (1%)	0 (0%)	0 (0%)	3 (8%)	0 (0%)	0.002	0 (0%)	1.000
Inotropes	142 (34%)	117 (35%)	4 (27%)	14 (36%)	3 (19%)	0.592	4 (33%)	1.000
Donor Characteristics								
Donor Age	38 ± 14	38 ± 14	44 ± 15	35 ± 15	44 ± 15	0.062	35 ± 16	0.418
Male Donor	274 (67%)	224 (68%)	8 (53%)	27 (71%)	9 (56%)	0.463	6 (50%)	0.211
Donor Sodium	144 ± 6	144 ± 6	145 ± 7	146 ± 4	143 ± 6	0.814	149 ± 7	0.060
Donor Troponin I	0.46 (0.19 - 0.88)	0.50 (0.20 - 0.90)	0.16 (0.12 - 0.20)	0.40 (0.17 - 1.47)	0.40 (0.34 - 0.44)	0.525	0.59 (0.22 - 1.00)	0.926
Dobutamine	1 (2%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)		0 (0%)	
Dopamine	1 (2%)	0 (0%)	1 (50%)	0 (0%)	0 (0%)		0 (0%)	
Epinephrine	1 (2%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)	0.073	0 (0%)	0.070
Levophed	3 (6%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)		2 (50%)	
Levothyroxine	41 (76%)	35 (82%)	1 (50%)	0 (0%)	3 (100%)		2 (50%)	
Vasopressin	7 (13%)	5 (12%)	0 (0%)	2 (100%)	0 (0%)		0 (0%)	
Donor EF	60 ± 7	60 ± 7	62 ± 8	57 ± 9	61 ± 4	0.060	61 ± 6	0.599
Left Ventricular Hypertrophy	21 (8%)	16 (8%)	1 (12%)	3 (12 %)	1 (12%)	0.491	0 (0%)	1.000
Transplant Characteristics								
Sex mismatch	104 (25%)	81 (24%)	6 (40%)	9 (23%)	5 (31%)	0.493	3 (25%)	1.000
Ischemic Time	3.7 ± 1.1	3.6 ± 1.1	3.5 ± 1.0	4.2 ± 1.1	4.4 ± 1.3	0.00005	3.8 ± 1.3	0.532
RADIAL score	2 (2 – 3)	2 (2 – 3)	3 (2 – 3)	3 (2 – 3)	3 (1 – 3)	0.2320	1 (2 – 3)	0.8331

Table 3 - Hemodynamic assessment from first 24 hours post-transplant

	All (n = 412)	None (n = 330)	Mild (n = 15)	Moderate (n = 39)	Severe (n = 16)	RV-PGD (n = 12)
LVEF ≤ 40%	43 (11%)	0 (0%)	9 (60%)	27 (75%)	7 (47%)	0 (0%)
RAP (mmHg)	13 ± 6	13 ± 6	17 ± 8	16 ± 7	17 ± 5	19 ± 3
Systolic PAP (mmHg)	33 ± 8	32 ± 8	39 ± 11	34 ± 9	29 ± 7	35 ± 6
Diastolic PAP (mmHg)	14 ± 7	14 ± 6	20 ± 6	16 ± 7	29 ± 9	5 ± 5
mean PAP (mmHg)	20 ± 6	20 ± 6	26 ± 7	22 ± 7	23 ± 8	16 ± 3
TPG (mmHg)	6 ± 3	6 ± 2	6 ± 2	6 ± 3	3 ± 2	10 ± 3
MAP (mmHg)	66 ± 17	66 ± 16	72 ± 5	62 ± 18	71 ± 13	52 ± 31
Cardiac Index (ml/min2)	2.2 ± 0.6	2.3 ± 0.6	1.8 ± 0.5	1.9 ± 0.4	1.4 ± 0.4	1.7 ± 0.1
Inotrope Score	17 (8 - 28)	16 (8 - 28)	8 (4 - 13)	27 (18 - 38)	26 (8 - 53)	24 (9 - 27)

Inotrope score = dopamine (X1) + dobutamine (X1) + amrinone (X1) + milrinone (X15) + epinephrine (X100) + norepinephrine (X100) with each drug dosed in µg/kg/min

Table 4 - Cox Proportional Hazard Model for impact of PGD on 1-year mortality, adjusted for age

Outcome	Study Results and measurements	Absolute effect estimates		Plain text summary
		Baseline	With predictor	
Mild LV-PGD	Hazard Ratio 2.40 (CI 95% 0.56 - 10.23) p-value 0.238	6 per 1000	14 per 1000	Mild LV-PGD makes little or no difference on 1-year all cause mortality
		Difference: 8 more per 1000 (CI 95% 2 less - 34 more)		
Modeate LV-PGD	Hazard Ratio 7.03 (CI 95% 3.39 - 14.59) p-value <0.001	6 per 1000	35 per 1000	Moderate LV-PGD increases the risk for 1-year all-cause mortality
		Difference: 29 more per 1000 (CI 95% 19 more - 59 more)		
Severe LV-PGD	Hazard Ratio 15.87 (CI 95% 7.20 - 34.98) p-value <0.001	6 per 1000	62 per 1000	Severe LV-PGD strongly increases the risk for 1-year all-cause mortallity
		Difference: 56 more per 1000 (CI 95% 36 more - 88 more)		
Recipient age 10-year increase	Hazard Ratio 0.75 (CI 95% 0.60 - 0.93) p-value 0.010	6 per 1000	5 per 1000	Increasing recipient age decreases the risk for 1-year all-cause mortality
		Difference: 1 fewer per 1000 (CI 95% 2 fewer - 0 fewer)		

Associations between each severity level of PGD and mortality is adjusted for recipient age. LV-PGD – left ventricular primary graft dysfunction.

Figure 1 – Rates of PGD over the years

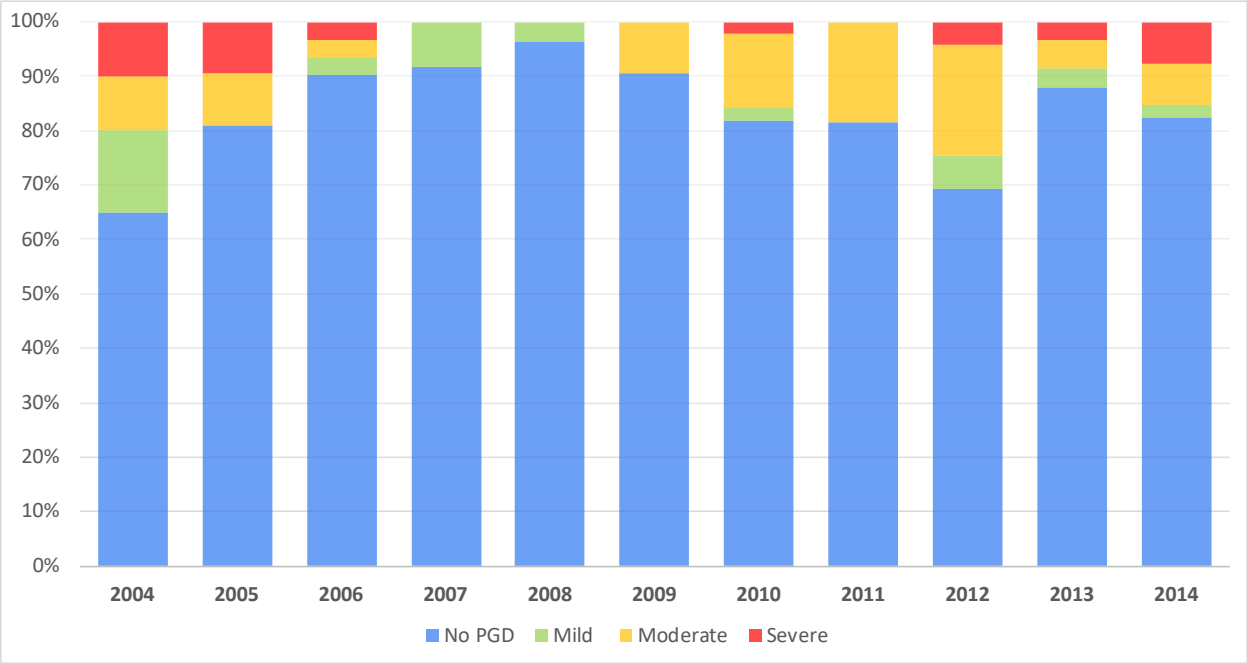
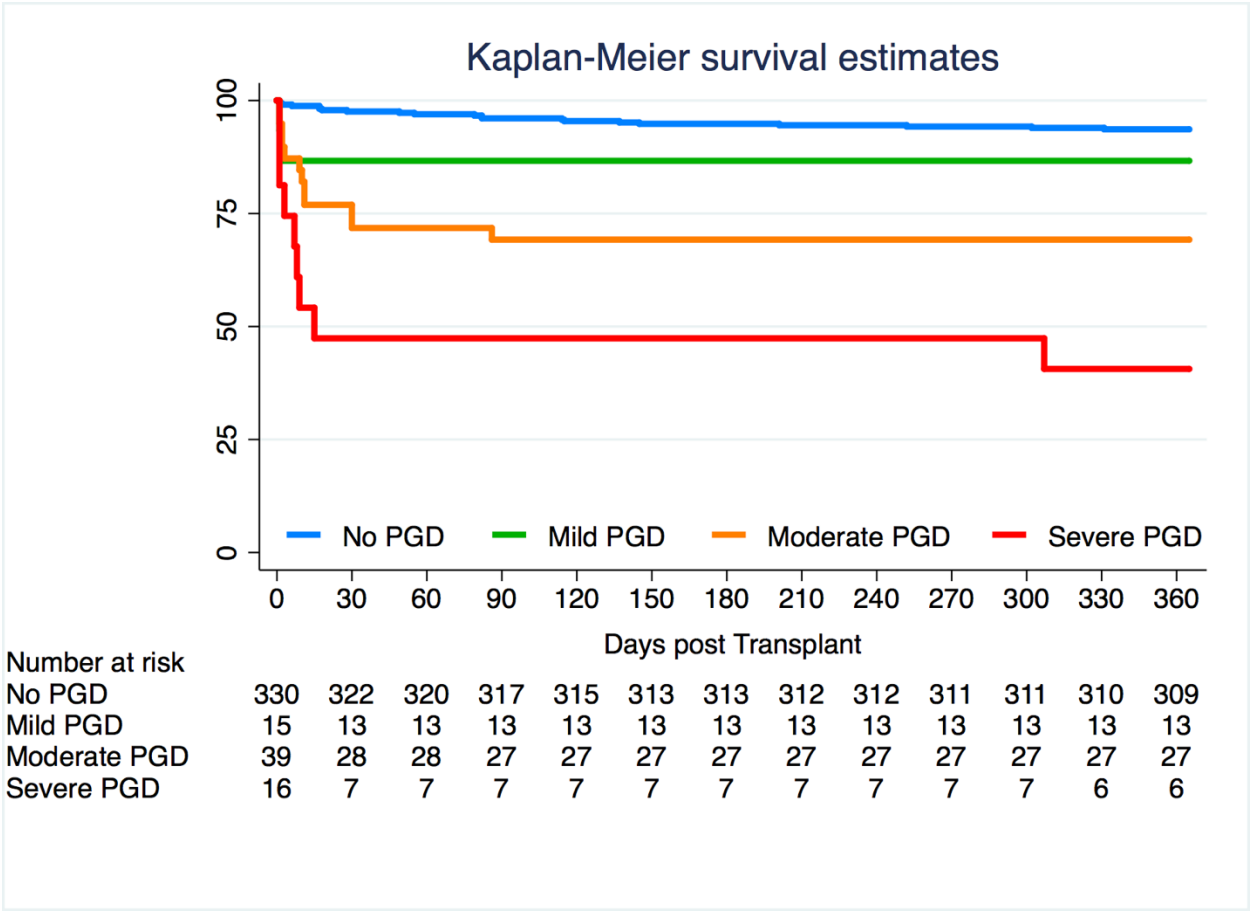


Figure 2 – Survival stratified by PGD severity



Number at risk													
No PGD	330	322	320	317	315	313	313	312	312	311	311	310	309
Mild PGD	15	13	13	13	13	13	13	13	13	13	13	13	13
Moderate PGD	39	28	28	27	27	27	27	27	27	27	27	27	27
Severe PGD	16	7	7	7	7	7	7	7	7	7	7	6	6

Chapter 3: Predictors of 1-year mortality in adult lung transplant recipients: protocol for a systematic review and meta-analysis.

STATUS: Published – BMC Systematic Reviews

https://doi.org/10.1186/s13643-019-1049-x

Farid Foroutan HBSc[1,2], Gordon Guyatt MSc MD[2], Erik Friesen HBSc[3], Luis Enrique Colunga Lozano MD[2], Aman Sidhu MSc, MD[1], Maureen Meade MD[2].

1- Multi-Organ Transplant, University Health Network, Toronto, Ontario, Canada

2- Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

3- University of Toronto, Toronto, Ontario, Canada

ABSTRACT

Background

Upon surviving the first-year post lung transplantation, recipients can expect a median survival of 8 years. Within the first year, graft failure and multi-organ failure (possibly secondary to graft failure) are common causes of mortality. To better understand prognosis within the first year, we plan on conducting a systematic review and meta-analysis of observational studies addressing the association between patient, donor, and transplant operative factors and graft loss 1-year post-lung transplant.

Methods

We searched Medline, Embase, Cochrane Database of Systematic Reviews, Cochrane Central Register, and PubMed supplemental for non-Medline records for observational studies identifying independent risk factors for early mortality (1-year) in adult lung transplant recipients. We plan on including cohort studies and secondary analyses of randomized controlled trials studying adult lung transplant recipients undergoing their first lung transplant, without any simultaneous organ transplant. We will conduct a random effects meta-analysis that pools effect estimates from all eligible studies to obtain a summary estimate and confidence interval for all independent non-therapeutic factors identified in the primary studies.

Discussion

The results from this study may inform future guidelines on the selection of candidates and donors for transplantation, predictive model development, and inform the decision-making process that the physician and patient undertake together. Furthermore, through the conduction of this review, we can identify limitations with the current best evidence, which will encourage the need for studies with better methodology to reassess predictors of mortality.

INTRODUCTION

Adult lung transplant recipients can expect a median survival of 8 years, conditional on surviving the first year post transplant[1]. Graft failure and multi-organ failure (possibly secondary to graft failure), however, are common causes of mortality within the first year, decreasing the conditional median survival from 8 years to 5 years[1].

To better understand the risk of mortality within the first year post transplantation, the International Society of Heart and Lung Transplantation (ISHLT) identified risk factors for 1-year mortality through the use of multivariable analysis[1]. Due to lack of mandatory reporting to the ISHLT registry, however, there may be studies from centers assessing the same set of risk factors but not reporting their findings to the registry. Based on their unique cohort, the risk estimates and 95% confidence interval calculated may vary compared to that reported by the registry. The risk estimates obtained from individual studies may also vary from one another which further necessitates the need for one true overall effect estimate with concomitant exploration of potential reasons for the observed inconsistency. For example, Borro et al. analyzed data from Spain and suggested an 8% relative reduction in hazard of mortality in patients undergoing double lung transplant compared to single lung transplant[2]. This association may have been observed by chance alone (95% CI crossing the boundary of no effect). Jacques et al. similarly analyzed Canadian data to suggest a 66% relative increase in hazard of mortality for patients undergoing bilateral lung transplantation (an association that again could be by chance alone)[3]. Finally, Neurohr et al. analyzed the same predictor (bilateral lung transplantation) to observe a 66% relative decrease in hazard of mortality (95% CI 0.14 to 0.79)[4]. A non-systematic narrative review and/or registry study suffers by only highlighting the one statistically significant risk

estimate obtained from a select sample of patients, and thus potentially ignoring the variability that may exist across different population.

A systematic review and meta-analysis addressing factors associated with mortality, identified by all individual studies, may better inform the relative independent effect of each risk factor in the context of patients suitable for transplantation. Among its advantages, by conducting a systematic review and meta-analysis, we may potentially generate hypotheses for sources of variations across studies assessing risk factors for early mortality. We therefore plan on conducting a systematic review and meta-analysis of observational studies addressing the association between patient, donor, and transplant operative factors and graft loss 1-year post-lung transplant.

RESEARCH QUESTION

In adult (≥18 years) lung transplant recipients, what are the independent predictors of mortality at 1-year post transplant?

OBJECTIVE

The objectives of this systematic review are:

- To summarize all factors associated with mortality within the first-year post lung transplant.
- To pool similar studies to obtain an overall point estimate and confidence interval of commonly identified risk factors for mortality at 1-year post transplant.

METHODS

We submitted our study protocol to the PROSPERO registry (submission ID: 132698). This study will systematically review observational studies identifying independent risk factors for early mortality (1-year) in adult lung transplant recipients. We will conduct a meta-analysis that pools effect estimates from all eligible studies to obtain a summary estimate and confidence interval for all independent factors identified in the primary studies.

Criteria for Study Selection

Types of studies

- Observational studies (retrospective or prospective) and secondary analyses of randomized control trials.

Types of Participants

- Adult (≥ 18 years) lung transplant recipients.

Type of exposure

- Any non-therapeutic independent predictor of mortality at 1-year post transplant related to recipient, donor, and the transplant operation.

Types of outcome measures

- The primary outcome measure is 1-year mortality post-transplant. Effect measures may be relative risk, odds ratio, or hazard ratio.

Language

- All languages

Selection process

We will include observational studies and secondary analyses of randomized control trials that enrolled adult (at least 95% of population = 18 years) de novo transplant recipients, evaluating any factor associated with mortality using multivariable analysis (Cox proportional hazards models, logistic regression models), reporting more than 20 events. We will exclude studies assessing mortality beyond the first year if they do not use a time to event analysis that meets the proportional hazard assumption. We do not plan on excluding studies based on language of publication.

We will exclude studies with insufficient information to generate estimates of effect for any predictor (lack of effect estimate, 95% confidence interval, p-value to be combined in the meta-analysis). We will exclude duplicate studies assessing the same population without additional data on new predictors. If two studies assessed the same population and predictors, we will include the study with larger sample size. If two studies, assessed the same population, but the smaller study reports on a unique predictor, the larger study will inform other predictors whereas the small study will inform the unique predictor.

Due to mandatory reporting to the United Network of Organ Sharing (UNOS) registry, we will consider risk estimates from this registry to represent all individual centers from the United States of America (USA) (thus excluding all individual studies from the USA for the same risk factor and time frame). Due to lack of mandatory reporting to the ISHLT registry, however, we will exclude all studies querying this registry. We will trust our search strategy and screening process to capture patients included in the ISHLT registry.

Using standardized study eligibility forms (Appendix B), paired reviewers will independently screen titles and abstracts of identified citations, and evaluate the full-text of articles deemed

potentially eligible by either reviewer. We will resolve disagreements in full-text screening through discussion or, if necessary, through adjudication by a third reviewer.

All citations will be imported into Covidence (Thomson Reuters)[5]. Ten reviewers (FF, EF, RZ, RK, AS, TW, AT, KM, SA, EA) will screen the titles and abstracts, as well as the full-text citations, independently and in duplicate. If a decision cannot be made on whether an article is eligible, reviewers will consult an independent adjudicator (GG).

Study sources

We completed the search using searched Medline, Embase, Cochrane Database of Systematic Reviews, Cochrane Central Register, and PubMed supplemental for non-Medline records. We developed a broad strategy for predictors/prognosis as many of the types of articles that might meet the inclusion criteria would not necessarily be found by a more focused search on predictors or the types of statistical tools used in the studies. That filter consists of a combination of Haynes' sensitive strategies for clinical prediction guides and etiology/risk, plus an added section related to prognostic factors and the statistical methods we are seeking.

Appendix A presents the complete search strategy. We restricted the search to adults, non-animal studies, non-conference proceedings/abstracts from Embase but without date or language restrictions.

Data abstraction

Data abstraction will be performed in duplicate by paired groups of reviewers. Standardized data collection forms will be created for reviewers to use when extracting data.

We will record the following data from each eligible article: author, year of publication, study type (retrospective, prospective, post-hoc analysis of trial), name of trial if post-hoc analysis of trial, whether single-center or multicenter, country of study, inclusion criteria, exclusion criteria, recruitment time frame (months), follow-up length (months), total sample size, definition of outcome (specific definition of mortality, and timing of outcome), number of events, baseline demographics of cohort (e.g sex, age, etiology of renal failure, body mass index) number of predictors included in the regression model, measure of point estimate of risk used (relative risk, odds ratio, hazard ratio), predictor, unit of change for continuous predictors, category for categorical predictors, measured point estimate of risk (odds ratio, hazard ratio), lower and upper confidence interval. Kaplan-Meier survivor function will provide estimated cumulative incidence of graft loss at 1-year post transplantation. If Kaplan-Meier curves are not available, data tables will provide the required data. All extracted data values will be rounded to the nearest integer.

Data synthesis

We plan on generating point estimates and respective 95% confidence intervals (CI) using hazard ratios (HRs). We will convert all estimates from each individual study to HR using baseline risk estimates from that study. If baseline risk estimates are not reported and conversion is not possible, we will conduct sub-group analyses to compare studies based on format of effect estimate (OR, RR, and HR). Specifically, we will compare studies presenting effect estimates using OR or RR, not reporting baseline risk (thus conversion not possible) with studies using HR together with studies reporting OR or RR but converted to HR due to presence of baseline risk. If no significant difference is observed, given the low risk of 1-year mortality, we will pool all risk

estimates and present as HR. In this sub-group analysis, we combined studies reporting in OR or RR together due to low risk of mortality in the first-year post transplant. We will pool HR estimates for each predictor through inverse variance analysis using random-effects meta-analysis. In studies with stratified groups, if we observed a linear association between the predictor and outcome, we will average the beta-coefficients across categories to obtain the estimate of effect associated with a unit change.

Certain publications may only report on significant predictors of mortality. Such studies raise problems of selective reporting bias as they do not report effect estimate and 95% confidence interval for variables fitted in the model, but not significantly associated with graft loss. For such studies, we will impute a relative effect estimate of 1, and utilize the hot deck approach for imputing the variance[6]. Subsequently, we will conduct a sensitivity analysis including the imputed studies.

We will consider a two-sided P value of 0.05 or less indicated statistical significance. Review Manager 5 will provide the software for statistical analyses, as well as forest plots and funnel plots.

Quality Assessment (Risk of bias within studies)

We will assess study risk of bias using a modified version of the Quality in Prognostic Studies (QUIPS) instrument[7] (Appendix C). We will assess study risk of bias using six domains (study participation, study attrition, prognostic factors, outcome measurement, study confounding, statistical analysis and reporting). We modified the QUIPS tool to rate each domain as low or high risk of bias as opposed to the original low, medium, and high risk of bias. Under the statistical analysis and reporting domain we modified QUIPS to assess if models are over fitted (less than

10 events for each binary prognostic factor). We will use the individual domains, rated as low, moderate, or high risk of bias, to inform the overall risk of bias in each study: five or six low-risk domains as overall low risk of bias, two or more high-risk domains as overall high risk of bias. Paired reviewers will independently assess each included study using the modified QUIPS tool.

Sources of Heterogeneity

We will address statistical heterogeneity through consistency of point estimates and extent of overlap of confidence interval. Heterogeneity will not be assessed with I2 statistics, as this is uniformly high and thus not useful in prognostic studies with very large sample size and precise estimates[8].

We will conduct subgroup analyses to explore heterogeneity across studies. We will focus on two possible effect modifiers: definition of mortality, duration of follow-up, and a number of aspects relating to risk of bias. We will conduct the following *a priori* subgroup analyses to explain heterogeneity:

1. Definition of mortality (comparing studies reporting on all-cause mortality, composite of mortality, re-transplantation, and graft loss, or composite of re-transplantation and graft loss.
2. We will include studies assessing mortality beyond the first year only if cox proportional hazard model is used to assess the independent impact of each prognostic factor. Therefore, we will conduct a sub-group analysis comparing the magnitude of effect estimate between studies specifically measuring mortality at 1-year to those looking at mortality beyond the first year. Certain prognostic factors may have a stronger association with graft loss at 1-year and thus inclusion

of follow-up time beyond the first year may attenuate the magnitude of effect estimate.

3. For each predictor, we will conduct sub-group analysis to compare effects in studies at high risk of bias with those at low risk of bias.

Confidence in estimate of effect

For assessing the overall confidence in certainty evidence, we will use the Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach considering risk of bias, imprecision, inconsistency, indirectness and publication bias[8]. Confidence could be rated as high, moderate, low, or very low. Because we are addressing the issue of prognosis and not causation, observational studies will start as high confidence. We will assess publication bias using funnel plots and visual inspection of symmetry. The GRADE approach is typically applied at the outcome level. Since the focus of our review is prognostic factors for the same outcome, assessment of confidence in certainty of evidence will be applied at the level of each individual prognostic factor.

If the pooled estimate of our sensitivity analysis (including imputed non-significant studies evaluating the predictor of interest) differs from our primary pooled estimate on the same predictor (not including the imputed non-significant predictors predictor), we will attribute more credibility and apply the GRADE assessment to the sensitivity analysis and rate down for risk of bias.

DISCUSSION

One operational challenge with our proposed review is the applicability of the GRADE tool for evaluating certainty in the evidence. Specifically, when assessing certainty, we need to address

presence of imprecision. Assessment of imprecision will require absolute effect estimates. To obtain absolute effect estimate, we will require baseline risk estimates obtained from patients who do not have the prognostic factor under evaluation. Not all included studies will require baseline risk estimates of graft survival in patients without the prognostic factor under evaluation. For this, we may need to refer to indirect evidence from registry studies to assess imprecision. This is a foreseeable issue that may arise with applicability of GRADE. The GRADE working group for prognosis is currently working on developing guidance for application of their tool when assessing the certainty of evidence for individual prognostic factors. We will conduct this review coincident with the development of their guidance which may ease our assessment for certainty of the evidence.

The results of this review may be beneficial for a number of future applications:

1) Transplant physicians and patients may be interested in accurate assessment of their prognosis. This review may provide information that physicians and patients can use in conversations regarding the extent of benefit they can expect from lung transplantation. If there are multiple prognostic factors that modify the risk of 1-year mortality, physicians with limited time may be challenged with combining the factors together to identify the specific risk of their patient. For this reason, risk prediction models may serve as an attractive tool in discussions of prognosis. To date, however, the performance of existing risk prediction models for 1-year mortality is just acceptable (AUC 0.67)[9]. Evidence from this review could inform the development of future risk prediction models that will result in improved prediction.

2) This review, will identify and summarize all predictors evaluated by individual studies. Small studies have access to more granular specific variables that may prove to be predictors of

1-year mortality. Public access to a repository of such studies will be a useful starting point for authors of future reviews that may be interested in evaluating these factors when more studies are available.

3) Finally, the current opioid crisis[10], efforts in expanding the pool of donors[11], and an ever-growing interest in better understanding the management of deceased donors[12] may change the risk of 1-year mortality. For this reason, it is useful better understand the risk of 1-year mortality by identifying and exploring characteristics of the low and high-risk lung transplant recipients.

ACKNOWLEDGEMENTS

The authors acknowledge Ms. Ani Orchania-Cheff for developing the search strategy for this review.

DISCLOSURES

None of the authors have any conflicts of interest to disclose.

REFERENCES

1. Yusen RD, Edwards LB, Dipchand AI, Goldfarb SB, Kucheryavaya AY, Levvey BJ, et al. The Registry of the International Society for Heart and Lung Transplantation: Thirty-third Adult Lung and Heart-Lung Transplant Report-2016; Focus Theme: Primary Diagnostic Indications for Transplant. J Heart Lung Transplant. 2016;35(10):1170-84.
2. Borro JM, Delgado M, Coll E, Pita S. Single-lung transplantation in emphysema: Retrospective study analyzing survival and waiting list mortality. World J Transplant. 2016;6(2):347-55.
3. Jacques F, El-Hamamsy I, Fortier A, Maltais S, Perrault LP, Liberman M, et al. Acute renal failure following lung transplantation: risk factors, mortality, and long-term consequences. Eur J Cardiothorac Surg. 2012;41(1):193-9.
4. Neurohr C, Huppmann P, Thum D, Leuschner W, von Wulffen W, Meis T, et al. Potential functional and survival benefit of double over single lung transplantation for selected patients with idiopathic pulmonary fibrosis. Transpl Int. 2010;23(9):887-96.
5. Abraham WT, Lindenfeld J, Reddy VY, Hasenfuss G, Kuck KH, Boscardin J, et al. A randomized controlled trial to evaluate the safety and efficacy of cardiac contractility modulation in patients with moderately reduced left ventricular ejection fraction and a narrow QRS duration: study rationale and design. Journal of cardiac failure. 2015;21(1):16-23.
6. Gelman A, Hill J. Missing-data imputation: New York: Cambridge University Press; 2006.
7. Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280-6.
8. Iorio A, Spencer FA, Falavigna M, Alba C, Lang E, Burnand B, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. Bmj. 2015;350:h870-h.
9. Russo MJ, Davies RR, Hong KN, Iribarne A, Kawut S, Bacchetta M, et al. Who is the high-risk recipient? Predicting mortality after lung transplantation using pretransplant risk factors. The Journal of thoracic and cardiovascular surgery. 2009;138(5):1234-8.e1.
10. Durand CM, Bowring MG, Thomas AG, Kucirka LM, Massie AB, Cameron A, et al. The Drug Overdose Epidemic and Deceased-Donor Transplantation in the United States: A National Registry Study. Ann Intern Med. 2018;168(10):702-11.
11. Moayedi Y, Fan CPS, Gulamhusein AF, Manlhiot C, Ross HJ, Teuteberg JJ, et al. Current Use of Hearts From Hepatitis C Viremic Donors. Circ Heart Fail. 2018;11(12):e005276.
12. Greer DM, Valenza F, Citerio G. Improving donor management and transplantation success: more research is needed. Intensive Care Med. 2015;41(3):537-40.

.

APPENDIX A – Search Strategy

Database: Ovid MEDLINE(R) <1946 to March Week 3 2017>
Search Strategy:
---

1 [lung transplantation] (0)
2 exp Lung Transplantation/ (13998)
3 exp lung/tr (105)
4 (lung? adj2 transplant*).mp. (17043)
5 (lung? adj2 graft*).mp. (579)
6 (lung? adj2 allograft*).mp. (1207)
7 (lung? adj2 allotransplant*).mp. (209)
8 (lung? adj2 heterograft*).mp. (0)
9 (lung? adj2 heterotransplant*).mp. (7)
10 (lung? adj2 homotransplant*).mp. (26)
11 (lung? adj2 homograft*).mp. (8)
12 (pulmonary adj2 transplant*).mp. (851)
13 (pulmonary adj2 graft*).mp. (287)
14 (pulmonary adj2 allograft*).mp. (289)
15 (pulmonary adj2 allotransplant*).mp. (15)
16 (pulmonary adj2 heterograft*).mp. (10)
17 (pulmonary adj2 heterotransplant*).mp. (2)
18 (pulmonary adj2 homotransplant*).mp. (13)
19 (pulmonary adj2 homograft*).mp. (383)
20 (cardiopulmonary adj2 transplant*).mp. (184)
21 (cardiopulmonary adj2 graft*).mp. (430)
22 (cardiopulmonary adj2 allograft*).mp. (3)
23 (cardiopulmonary adj2 allotransplant*).mp. (2)
24 (cardiopulmonary adj2 heterograft*).mp. (0)
25 (cardiopulmonary adj2 heterotransplant*).mp. (0)
26 (cardiopulmonary adj2 homotransplant*).mp. (2)
27 (cardiopulmonary adj2 homograft*).mp. (2)
28 or/2-27 (18727)
29 [graft loss etc] (0)
30 exp Host vs Graft Reaction/ (86666)
31 (graft? adj2 loss*).mp. (6005)
32 (transplant adj2 loss*).mp. (214)
33 Primary Graft Dysfunction/ (480)
34 (graft? adj2 dysfunction*).mp. (3006)
35 (transplant adj2 dysfunction*).mp. (406)
36 (graft? adj2 fail*).mp. (10103)
37 (transplant adj2 fail*).mp. (979)
38 (graft? adj2 survival*).mp. (49822)
39 (transplant adj2 survival*).mp. (2126)

40 exp mortality/ (330045)
41 mo.fs. (494027)
42 or/30-41 (759856)
43 28 and 42 (6711)
44 [predictors/risk/prognostic factors] (0)
45 predict*.mp. (1116690)
46 scor*.tw. (629282)
47 observ*.mp. (2638814)
48 validat*.mp. (353136)
49 exp risk/ (1007713)
50 risk*.mp. (1892186)
51 exp Cohort Studies/ (1650322)
52 between group*.tw. (85712)
53 exp prognosis/ (1352527)
54 (prognos* adj2 factor*).mp. (75370)
55 (prognos* adj2 value*).mp. (36679)
56 (associat* adj2 factor*).mp. (121225)
57 (independent adj2 factor*).mp. (58971)
58 (multivariate adj2 factor*).mp. (3166)
59 (multivariable* adj2 factor*).mp. (513)
60 exp Regression Analysis/ (360244)
61 regression*.mp. (549354)
62 (hazard* adj2 model*).mp. (76125)
63 (cox adj2 model*).mp. (14721)
64 (hazard* adj2 ratio*).mp. (67230)
65 exp survival analysis/ (236746)
66 or/45-65 (7000435)
67 43 and 66 (4447)
68 animals/ not (animals/ and humans/) (4329339)
69 67 not 68 (4160)
70 limit 69 to ("all infant (birth to 23 months)" or "all child (0 to 18 years)") (1178)
71 limit 69 to "all adult (19 plus years)" (2935)
72 69 not 70 (2982)
73 71 or 72 (3863)
74 remove duplicates from 73 (3781)

Database: Ovid MEDLINE(R) Epub Ahead of Print and In-Process & Other Non-Indexed Citations <March 23, 2017>
Search Strategy:
---
1 (lung? adj2 transplant*).mp. (1298)
2 (lung? adj2 graft*).mp. (43)
3 (lung? adj2 allograft*).mp. (116)
4 (lung? adj2 allotransplant*).mp. (2)

5 (lung? adj2 heterograft*).mp. (0)
6 (lung? adj2 heterotransplant*).mp. (0)
7 (lung? adj2 homotransplant*).mp. (0)
8 (lung? adj2 homograft*).mp. (0)
9 (pulmonary adj2 transplant*).mp. (63)
10 (pulmonary adj2 graft*).mp. (33)
11 (pulmonary adj2 allograft*).mp. (13)
12 (pulmonary adj2 allotransplant*).mp. (0)
13 (pulmonary adj2 heterograft*).mp. (0)
14 (pulmonary adj2 heterotransplant*).mp. (0)
15 (pulmonary adj2 homotransplant*).mp. (0)
16 (pulmonary adj2 homograft*).mp. (20)
17 (cardiopulmonary adj2 transplant*).mp. (14)
18 (cardiopulmonary adj2 graft*).mp. (22)
19 (cardiopulmonary adj2 allograft*).mp. (0)
20 (cardiopulmonary adj2 allotransplant*).mp. (1)
21 (cardiopulmonary adj2 heterograft*).mp. (0)
22 (cardiopulmonary adj2 heterotransplant*).mp. (0)
23 (cardiopulmonary adj2 homotransplant*).mp. (0)
24 (cardiopulmonary adj2 homograft*).mp. (0)
25 or/1-24 (1429)
26 (graft? adj2 loss*).mp. (645)
27 (transplant adj2 loss*).mp. (22)
28 (graft? adj2 dysfunction*).mp. (319)
29 (transplant adj2 dysfunction*).mp. (25)
30 (graft? adj2 fail*).mp. (1004)
31 (transplant adj2 fail*).mp. (109)
32 (graft? adj2 survival*).mp. (1414)
33 (transplant adj2 survival*).mp. (306)
34 mortal*.mp. (74315)
35 or/26-34 (77014)
36 25 and 35 (388)

Database: Embase <1974 to 2017 March 23>
Search Strategy:
---
1 exp lung transplantation/ (31300)
2 (lung? adj2 transplant*).mp. (36600)
3 (lung? adj2 graft*).mp. (2621)
4 (lung? adj2 allograft*).mp. (2092)
5 (lung? adj2 allotransplant*).mp. (258)
6 (lung? adj2 heterograft*).mp. (0)
7 (lung? adj2 heterotransplant*).mp. (9)
8 (lung? adj2 homotransplant*).mp. (23)

9 (lung? adj2 homograft*).mp. (8)
10 (pulmonary adj2 transplant*).mp. (1369)
11 (pulmonary adj2 graft*).mp. (462)
12 (pulmonary adj2 allograft*).mp. (385)
13 (pulmonary adj2 allotransplant*).mp. (20)
14 (pulmonary adj2 heterograft*).mp. (10)
15 (pulmonary adj2 heterotransplant*).mp. (1)
16 (pulmonary adj2 homotransplant*).mp. (12)
17 (pulmonary adj2 homograft*).mp. (564)
18 (cardiopulmonary adj2 transplant*).mp. (263)
19 (cardiopulmonary adj2 graft*).mp. (528)
20 (cardiopulmonary adj2 allograft*).mp. (3)
21 (cardiopulmonary adj2 allotransplant*).mp. (3)
22 (cardiopulmonary adj2 heterograft*).mp. (0)
23 (cardiopulmonary adj2 heterotransplant*).mp. (0)
24 (cardiopulmonary adj2 homotransplant*).mp. (2)
25 (cardiopulmonary adj2 homograft*).mp. (3)
26 or/1-25 (39025)
27 graft dysfunction/ (5238)
Annotation: Did not explode but chose the relevant narrower terms.
28 delayed graft function/ (4711)
29 graft failure/ (31128)
30 primary graft dysfunction/ (1211)
31 lung graft rejection/ (1741)
32 (graft? adj2 loss*).mp. (12152)
33 (transplant adj2 loss*).mp. (425)
34 (graft? adj2 dysfunction*).mp. (9442)
35 (transplant adj2 dysfunction*).mp. (655)
36 (graft? adj2 fail*).mp. (36385)
37 (transplant adj2 fail*).mp. (1873)
38 (graft? adj2 survival*).mp. (63703)
39 (transplant adj2 survival*).mp. (4652)
40 exp mortality/ (971440)
41 or/27-40 (1062643)
42 26 and 41 (11870)
43 predict*.tw. (1588973)
44 exp methodology/ (5081887)
45 validat*.tw. (525071)
46 risk*.mp. (3065038)
47 exp epidemiology/ (2941671)
48 prognosis/ (597088)
49 prognostic assessment/ (3274)
50 (prognos* adj2 factor*).mp. (125881)
51 (prognos* adj2 value*).mp. (61042)

52 (associat* adj2 factor*).mp. (189556)
53 (independent adj2 factor*).mp. (98764)
54 (multivariate adj2 factor*).mp. (5304)
55 (multivariable* adj2 factor*).mp. (956)
56 exp regression analysis/ (541847)
57 regression*.mp. (854778)
58 (hazard* adj2 model*).mp. (126317)
59 (cox adj2 model*).mp. (29455)
60 (hazard* adj2 ratio*).mp. (115817)
61 survival analysis/ (3722)
62 (survival adj2 analy*).mp. (56984)
63 or/43-62 (9944135)
64 42 and 63 (9606)
65 (exp animals/ or exp animal experimentation/ or nonhuman/) not ((exp animals/ or exp animal experimentation/ or nonhuman/) and exp human/) (5969916)
66 64 not 65 (9362)
67 limit 66 to (embryo <first trimester> or infant <to one year> or child <unspecified age> or preschool child <1 to 6 years> or school child <7 to 12 years> or adolescent <13 to 17 years>) (1338)
68 limit 66 to (adult <18 to 64 years> or aged <65+ years>) (3671)
69 66 not 67 (8024)
70 68 or 69 (8824)
71 limit 70 to (book or book series or chapter or conference abstract or conference paper or conference proceeding or "conference review") (3729)
72 70 not 71 (5095)
73 remove duplicates from 72 (4820)

Database: EBM Reviews - Cochrane Database of Systematic Reviews <2005 to March 22, 2017>
Search Strategy:
---
1 (lung? adj2 transplant*).mp. (52)
2 (lung? adj2 graft*).mp. (4)
3 (lung? adj2 allograft*).mp. (1)
4 (lung? adj2 allotransplant*).mp. (0)
5 (lung? adj2 heterograft*).mp. (0)
6 (lung? adj2 heterotransplant*).mp. (0)
7 (lung? adj2 homotransplant*).mp. (0)
8 (lung? adj2 homograft*).mp. (0)
9 (pulmonary adj2 transplant*).mp. (2)
10 (pulmonary adj2 graft*).mp. (1)
11 (pulmonary adj2 allograft*).mp. (0)
12 (pulmonary adj2 allotransplant*).mp. (0)
13 (pulmonary adj2 heterograft*).mp. (0)
14 (pulmonary adj2 heterotransplant*).mp. (0)

15 (pulmonary adj2 homotransplant*).mp. (0)
16 (pulmonary adj2 homograft*).mp. (0)
17 (cardiopulmonary adj2 transplant*).mp. (0)
18 (cardiopulmonary adj2 graft*).mp. (1)
19 (cardiopulmonary adj2 allograft*).mp. (0)
20 (cardiopulmonary adj2 allotransplant*).mp. (0)
21 (cardiopulmonary adj2 heterograft*).mp. (0)
22 (cardiopulmonary adj2 heterotransplant*).mp. (0)
23 (cardiopulmonary adj2 homotransplant*).mp. (0)
24 (cardiopulmonary adj2 homograft*).mp. (0)
25 or/1-24 (56)
26 (graft? adj2 loss*).mp. (62)
27 (transplant adj2 loss*).mp. (6)
28 (graft? adj2 dysfunction*).mp. (7)
29 (transplant adj2 dysfunction*).mp. (1)
30 (graft? adj2 fail*).mp. (79)
31 (transplant adj2 fail*).mp. (13)
32 (graft? adj2 survival*).mp. (61)
33 (transplant adj2 survival*).mp. (18)
34 mortal*.mp. (4689)
35 or/26-34 (4724)
36 25 and 35 (47)

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <February 2017>
Search Strategy:
---
1 exp Lung Transplantation/ (186)
2 exp lung/tr (0)
3 (lung? adj2 transplant*).mp. (601)
4 (lung? adj2 graft*).mp. (66)
5 (lung? adj2 allograft*).mp. (27)
6 (lung? adj2 allotransplant*).mp. (0)
7 (lung? adj2 heterograft*).mp. (0)
8 (lung? adj2 heterotransplant*).mp. (0)
9 (lung? adj2 homotransplant*).mp. (0)
10 (lung? adj2 homograft*).mp. (0)
11 (pulmonary adj2 transplant*).mp. (23)
12 (pulmonary adj2 graft*).mp. (14)
13 (pulmonary adj2 allograft*).mp. (1)
14 (pulmonary adj2 allotransplant*).mp. (0)
15 (pulmonary adj2 heterograft*).mp. (0)
16 (pulmonary adj2 heterotransplant*).mp. (0)
17 (pulmonary adj2 homotransplant*).mp. (0)
18 (pulmonary adj2 homograft*).mp. (8)

19 (cardiopulmonary adj2 transplant*).mp. (4)
20 (cardiopulmonary adj2 graft*).mp. (191)
21 (cardiopulmonary adj2 allograft*).mp. (1)
22 (cardiopulmonary adj2 allotransplant*).mp. (0)
23 (cardiopulmonary adj2 heterograft*).mp. (0)
24 (cardiopulmonary adj2 heterotransplant*).mp. (0)
25 (cardiopulmonary adj2 homotransplant*).mp. (0)
26 (cardiopulmonary adj2 homograft*).mp. (0)
27 or/1-26 (850)
28 exp Host vs Graft Reaction/ (2758)
29 (graft? adj2 loss*).mp. (809)
30 (transplant adj2 loss*).mp. (39)
31 Primary Graft Dysfunction/ (13)
32 (graft? adj2 dysfunction*).mp. (218)
33 (transplant adj2 dysfunction*).mp. (36)
34 (graft? adj2 fail*).mp. (1027)
35 (transplant adj2 fail*).mp. (72)
36 (graft? adj2 survival*).mp. (2641)
37 (transplant adj2 survival*).mp. (222)
38 exp mortality/ (10980)
39 mo.fs. (21900)
40 or/28-39 (29512)
41 27 and 40 (193)

PubMed supplemental for non-Medline records

Query	**Items found**
Search ((((((("lung"[All Fields] OR "lungs"[All Fields] OR "pulmonary"[All Fields] OR "cardiopulmonary"[All Fields])) AND ("transplant"[All Fields] OR "transplants"[All Fields] OR "transplantation"[All Fields] OR "transplanted"[All Fields] OR "graft"[All Fields] OR "grafts"[All Fields] OR allograft*[All Fields] OR allotransplant*[All Fields] OR heterograft*[All Fields] OR heterotransplant*[All Fields] OR homotransplant*[All Fields] OR homograft*[All Fields]))) AND (((((("graft"[All Fields] OR "grafts"[All Fields] OR "transplant"[All Fields])) AND (loss*[All Fields] OR dysfunction*[All Fields] OR fail*[All Fields] OR survival*[All Fields]))) OR mortal*[All Fields]))) AND ((((((((((predict*[Title/Abstract] OR scor*[Title/Abstract] OR observ*[Title/Abstract] OR "validation"[Title/Abstract] OR "validate"[Title/Abstract] OR risk*[Title/Abstract] OR "cohort"[Title/Abstract])) OR group*[Text Word]) OR regression*[Text Word]) OR (hazard*[Text Word] AND model*[Text Word])) OR ("cox"[Text Word] AND model*[Text Word])) OR (hazard*[Text Word] AND ratio*[Text Word]))) OR (((prognos*[Text Word] OR associat*[Text Word] OR "independent"[Text Word] OR "multivariate"[Text Word] OR multivariable*[Text Word])) AND (factor*[Text Word]	586

Query	Items found
OR variable*[Text Word]))))) OR (survival[Text Word] AND analy*[Text Word]))))) AND (pubstatusaheadofprint OR publisher[sb] OR pubmednotmedline[sb])	

APPENDIX B – Study Eligibility Form

Population:		
• Lung Transplant Recipients	**YES**	**NO**
• Adults (≥ 18 years old)	**YES**	**NO**

Predictor:		
• Any predictor of mortality (see outcome details below)	**YES**	**NO**

Adjusted Analysis:		
• Multivariate analysis	**YES**	**NO**

Outcomes reported:		
• Mortality within the first-year post transplant OR Graft loss at any time post-transplant ONLY IF authors used Cox regression analysis (Hazard Ratio, HR)	**YES**	**NO**

Type of article:		
• Cohort study (retrospective or prospective) or • RCT cohort (post-hoc analysis)	**YES**	**NO**

Duplicated population:		
• If duplicated, does this study provide new information? • If duplicated, is study more recent?	**YES**	**NO**

Study inclusion:	
• All the answers are YES	**INCLUDE**
• Any answer is NO	**EXCLUDE**
• If you are unsure of the answer, include for full text screening	**INCLUDE**

Instructions:

[1] On occasion, some of the above criteria, especially during T&A screening, will be unclear. If any response to the above questions is UNCLEAR, mark YES.

[2] Consider YES if any type of predictor, including but not limited to clinical characteristics, laboratory values, test results and any other clinical event

[3]Exclude studies evaluating therapies as a predictor and not reporting on any other potential predictor.

[5] Consider NO if the study used only any other type of adjustment for potential confounders, including matched design or stratification.

[6] Outcomes could be analyzed at any time point during follow up if hazard ratio is reported. 1-year mortality should be present for studies using odds ratio.

[7] Consider YES if it is a post-hoc analysis of an RCT evaluating other predictors and not just a therapy. Meta-analysis on observational studies evaluating a therapy could be included in the individual studies used

multivariable analysis. We will abstract information related to predictors from individual studies if they fulfill the rest of this project inclusion criteria

APPENDIX C – Risk of Bias Assessment Tool

Modified version of QUIPS

1. Study Participation. Adequately described:
    - Source of target population.
    - Sampling frame and recruitment method.
    - Period of recruitment.
    - Place of recruitment.
    - Inclusion/exclusion criteria
    - Baseline key characteristics

2. Study Attrition. Adequately described:
    - Lost to follow-up.
    - Attempts at collecting information on patients lost to follow-up.
    - Reasons for lost to follow-up.
    - Key characteristics of participants lost to follow-up.
    - No differences between patients who completed and those lost to follow-up

3. Prognostic Factors. Adequately described:
    - Definition of prognostic factors (if definition required).
    - Methods of prognostic factor measurement (if measurement required).
    - Adequate proportion of the study sample has complete data on prognostic factor.
    - Methods of imputation for dealing with missing prognostic factor data.

4. Outcome. Adequately described:
    - Definition of outcome (including duration of follow-up).
    - Methods of outcome measurement (if measurement required).
    - Methods and setting of outcome measurement are the same for all study participants.

5. Study Confounding. Adequately described:
    - All important confounders, including treatments.
    - Important confounders are accounted for in the final regression model.

6. Statistical Analysis and Reporting. Adequately described:
    - Analytic strategy.
    - Model building (low risk of bias for building model based on conceptual framework.
    - No overfitting of final model (1 variable for every 10 events).
    - Checking of model assumptions.
    - Reporting on all variables included in the final model (statistically significant or not).

Chapter 4: Predictors of 1-year mortality post Lung Transplantation: Systematic review and Meta-analysis

STATUS: Unpublished

Farid Foroutan[1,3], Abdullah Malik[1], Kathryn Elizabeth Clark[1], Tayler A. Buchan[1], Caitlin A. Elmslie, Erik Loewen Friesen[2], Aemal Akhtar[1], Alanna Rigobon[2], Madeleine Stein[1], Juan Jose Yepes Nunez[3], Daniella C. Birriel[1], Aman Sidhu[1], Gordon Guyatt[3], Maureen Meade[3]

1- Ted Rogers Centre for Heart Research, Multi-Organ Transplant Program, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

2- Department of Medicine, University of Toronto, Toronto, Ontario, Canada

3- Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

ABSTRACT

Background

In lung transplantation, prognostic factors may help in identifying characteristics of recipients and donors that influence allograft and patient survival. Understanding such factors can guide clinical decision making and better matching of recipients and donors. Through systematic review and meta-analysis, we aimed to provide best estimates for the association between characteristics of donors, transplant operation, recipients, or post-transplant complications, and the risk of 1-year all-cause mortality in adult lung transplant recipients.

Methods

We systematically searched 5 bibliographic databases for eligible primary studies assessing the association between any potential risk factor (related to lung donor, recipient, or the transplant operation), and 1-year recipient mortality. We pooled effect estimates from the primary studies using a random effects or fixed effects framework, according to a pre-specified protocol. This review utilized the GRADE approach to assess the quality of the evidence.

Results

With high certainty in the evidence, we identified several significant predictors of 1-year mortality: donor sex (HR 0.92 for male vs female, 95% CI 0.88 – 0.98), type of transplant (HR 0.81 for bilateral vs single lung transplant, 95% CI 0.75 – 0.87), CMV mismatch (HR 1.26 for negative recipient and positive donor, 95% CI 1.11 – 1.44), cardiopulmonary bypass use (HR 1.31, 95% CI 1.03 – 1.68), recipient age less than 20 years (HR 1.37 for age <20 vs ≥20, 95% CI 1.16 – 1.60), otherwise older recipient age (HR 1.31 per 10-year increase, 95% CI 1.03 – 1.68), recipient hypertension (HR 1.34, 95% CI 1.04 – 1.73), pre-transplant lung disease (HR 0.85 for obstructive vs restrictive disease, 95% CI 0.78 – 0.92), pre-transplant coronary artery disease

(HR 1.58, 95% CI 1.13 – 2.22), post-transplant severe primary graft dysfunction (PGD 3) (1.66, 95% CI 1.09 – 2.51), re-exploration for bleeding (HR 1.22, 95% CI 1.10 – 1.36), post-op need for ECMO (HR 1.91, 95% CI .179 – 2.04), need for vasopressin (HR 2.12, 95% CI 1.79 – 4.13), need for dialysis (HR 7.87, 95% CI 6.79 – 9.12), and cardiac complications (HR 1.25, 95% CI 1.13 – 1.39).

Conclusion

With a high degree of certainty, we identified 17 prognostic factors that impact the risk of 1-year mortality. With the exception of post-transplant dialysis, however, the impact of each factor is modest, suggesting the potential value for a predictive model that would incorporate multiple factors. This in turn may guide clinicians in identifying the highest risk recipients and donors.

INTRODUCTION

Adults undergoing lung transplantation can anticipate a median survival of 8-years[1]. This prognosis, however, is conditional on surviving the first-year post lung transplantation. During the first year, lung recipients are most likely to die from graft failure, multi-organ failure (possibly secondary to graft failure), or infections[1]. These in turn reduce the median anticipated survival from 8 to 6 years[1].

Large transplant registries annually analyze and report the associations of recipient, donor, and transplant operation variables with the risk of 1-year mortality[1]. This helps to elucidate characteristics associated with higher or lower risk of 1-year mortality. Although these reports provide rapid updates on changing trends in lung transplantation, challenges inherent to their design decrease the certainty in their results[2]. For example, eligible patients are typically omitted due to uninterpretable, outlier, or missing data related to the relevant predictors or outcomes. The extent of bias in registry reports is often unclear due to limited reporting of analytic procedures.

Reports from individual centers typically provide information that allows an assessment of the potential bias in their findings regarding risk factors for specific transplant outcomes[3]. The measured effect of a given prognostic factor may vary between studies, however, necessitating exploration possible reasons for the observed inconsistencies. For example, Borro et al. found an 8% relative reduction (95% CI 53% relative reduction to 80% relative increase) in the hazard of mortality for patients undergoing bilateral lung transplant (BLTx) compared to single lung transplant (SLTx) for emphysema[4]. Similarly, Jacques et al. analyzed Canadian data and showed a 66% relative increase in the hazard of mortality (95% CI 31% relative reduction to 4 fold relative increase) for patients undergoing BLTx[5]. Neurohr et al. analyzed the same predictor (BLTx) among

German transplant recipients and observed a 66% relative decrease in the hazard of mortality (95% CI 0.14 – 0.79)[5]. The wide confidence limits reflect the small sample sizes and resultant imprecision in these study estimates.

A systematic review and meta-analysis of prognostic factors for mortality will better inform the relative effect of each risk factor [6]. To this end, we conducted a systematic review and meta-analysis of observational studies addressing the association between patient, donor, and transplant operative factors with 1-year mortality post-lung transplant.

METHODS

We previously published a detailed protocol for this systematic review,[7] and registered the review with PROSPERO(CRD42019132698). A summary of the review methods follows.

Data sources and searches

An experienced information specialist (AOC) developed a broad search strategy for predictors and prognosis in lung transplantation available in the supplemental material. Using this strategy, in March 2017 we searched Medline, EMBASE, Cochrane Database of Systematic Reviews, Cochrane Central Register and PubMed supplemental for non-Medline records.

Study selection

We included observational studies and secondary analyses of randomized controlled trials that enrolled adult lung transplant recipients, and evaluated any prognostic factor using multivariable regression analyses (Cox proportional hazard models and logistic regression), and reported more than 20 events. We excluded studies including re-transplantation in their cohorts. We did not restrict inclusion based on the language of publication, nor did we select against studies based

on whether they reported on significant or non-significant associations between evaluated prognostic factors and the outcome of interest.

We excluded studies that evaluated mortality beyond the first year and did not use time-to-event analysis and studies with insufficient information to generate effect estimates and 95% confidence interval for any predictor. When studies reported more than once on the same population and predictors, we included the study with the larger sample size. When two studies reported on the same population but the smaller study reported on a unique predictor, the larger study informed our assessment of the common predictors whereas the smaller study informed the unique predictor.

Due to mandatory reporting, by each transplant center, to the United Network of Organ Sharing (UNOS) registry, we considered associations identified from this registry to represent all individual centers from the United States of America (USA). By doing so, we excluded all individual studies from the USA for the same risk factor and same enrollment period. Due to lack of mandatory reporting to the International Society of Heart and Lung Transplantation (ISHLT) registry, however, we excluded any study utilizing this resource.

Working in pairs, reviewers independently screened titles and abstracts of identified citations, and evaluated the full text of articles deemed potentially eligible by either reviewer. We resolved disagreements through discussion or adjudication by a third reviewer (FF).

Data abstraction

Reviewers, working independently and in duplicate, extracted from eligible studies key information related to the center(s), recruitment period, definition and number of primary outcomes. We also extracted information related to the characteristics of the cohorts. Reviewers

also abstracted key information regrding each predictor: the definition, estimate of effect, confidence interval and the covariates adjusted for in the final regression model.

Risk of bias of individual studies

We assessed the risk of bias of individual studies using the Quality in Prognostic Studies (QUIPS) instrument[8]. When we judged at least 5 the 6 QUIPS domains to be at low risk of bias, we classified the overall risk of bias as low; otherwise we considered studies at high risk of bias.

Data synthesis and statistical analysis

The primary studies always reported point estimates and 95% confidence intervals (CIs) as hazard ratios (HR), odds ratios (OR) or relative risks (RR). Due to the low risk of 1-year mortality, we included OR and RR in the same meta-analysis without conversion[9 10]. To combine studies that reported HR with those reporting OR or RR, we conducted sub-group comparisons. When we observed a clinically or statistically significant difference between binary (e.g., OR or RR) and time-to-event measures (e.g., HR), we converted the OR or RR to HR using baseline risk estimates from the individual studies. When primary study reports did not provide baseline risk estimates, we utilized the average risk, prevalence of the prognostic factor, and the relative effect to estimate the baseline risk[11].

We addressed statistical heterogeneity through visual inspection of forest plots, looking for the consistency of point estimates and the extent of overlap in confidence intervals. We did not assess heterogeneity using the I^2 statistic, since this approach is not useful in prognostic studies with a very large sample size and precise estimates[12].

This review focused on two possible subgroup analyses: risk of bias and outcome definition. For definition of mortality, we compared studies reporting on all-cause mortality to those reporting a composite of mortality, re-transplantation, and graft loss, or composite of re-

transplantation and graft loss. We hypothesized that donor and transplant factors may present stronger associations with graft-loss or re-transplantation, whereas recipient factors may show stronger associations with all-cause mortality. Furthermore, we conducted a sub-group analysis comparing the magnitude of effect estimate between studies specifically measuring mortality at 1-year to those looking at mortality beyond the first year. Certain prognostic factors may have a stronger association with mortality at 1-year and thus inclusion of follow-up time beyond the first year may attenuate the magnitude of effect estimate.

For each predictor, we conducted a sub-group analysis to compare effects in studies at high risk of bias with those at low risk of bias, with the hypothesis that studies at high risk of bias observed stronger associations between the prognostic factors and the outcome.

When the sub-group analysis showed a significant difference across groups, we focused the analysis on studies at low risk of bias and/or those assessing outcomes using the same definition and applied the GRADE assessment only on these studies. We applied a two-sided P value of 0.05 or less to denote statistical significance. STATA's *metan* function provided the platform for conducting all statistical analyses[13].

Certainty in the body of evidence

To assess the certainty of evidence across all studies related to a given predictor, we used The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) approach that rates the certainty of evidence as high, moderate, low, or very low considering issues of risk of bias in the primary studies, imprecision in the final measures of association, inconsistency in results across studies, indirectness and publication bias[14]. We assessed for publication bias using visual inspection of funnel plots.

RESULTS

Study selection and characteristics

The literature search identified 16,098 citations. After screening titles and abstracts, reviewers judged 516 articles eligible for full text review. Among these, only 47 met criteria for inclusion in this review[4 5 15-59]. Figure 1 summarizes the reasons for study exclusion during the full text review. The supplemental material contains a table summarizing the characteristics of the included studies and their cohorts.

Risk of bias of individual studies

Of the 47 included studies, reviewers judged 13 to be at high risk of bias[5 16 20 27 28 33 36 37 42 46 48 56-58]. The most common sources of bias related to statistical analyses and reporting (over fitting of the regression models, building a multivariable model based on level of significance in univariable analysis, and inclusion of collinear variables such BMI, weight and height). Study authors generally developed multivariable regression models including 12 variables (standard deviation of 9, minimum of 2 and maximum of 36). We found no relation between risk of bias and magnitude of effect.

Meta-analyses of donor factors

This review identified 3 donor factors amenable to a meta-analysis. Only 1 donor factor, male sex, was associated with a lower risk of 1-year mortality beyond chance (HR 0.92, 95% CI 0.88 – 0.97; high certainty in effect estimate) (table 1). With moderate certainty, We found no relation between risk of 1-year mortality with increasing donor age (per 10-year increase), or history of smoking (moderate certainty (table 1).

Meta-analyses of transplant process factors

We analyzed 9 variables associated with the transplant process (table 1). With moderate to high certainty, we observed associations between the following prognostic factors and the risk

of 1-year mortality: type of transplant (HR 0.81 for BLTx vs SLTx, 95% CI 0.75 – 0.87; high certainty), CMV mismatch (HR 1.14, 95% CI 1.07 – 1.22; high certainty), CMV status (HR 1.26 for recipient negative and donor positive, 95% CI 1.11 – 1.44; moderate certainty), CMV status (HR 1.09 for recipient positive and donor positive, 95% CI 1.02 – 1.17; high certainty), and use of CPB (HR 1.31, 95% CI 1.03 – 1.68). With moderate to high certainty, we excluded a statistically significant associations for ischemic time (per 1-hour increase), and CMV status (recipient positive and donor negative).

Meta-analyses of recipient factors

With high certainty, we observed associations with 1-year mortality for recipient age (HR 1.16 per 10-year increase, 95% CI 1.13 – 1.19), obstructive lung disease compared to restrictive (HR 0.85, 95% CI 0.78 – 0.92), and history of coronary artery disease (HR 1.58, 95% CI 1.13 – 2.22). With moderate certainty, we found a likely increase in the hazard of mortality associated with hypertension (HR 1.34, 95% CI 1.04 – 1.73) and underweight BMI (HR 1.30 for BMI <18 vs 18 – 23, 95% CI 1.14 – 1.49). Although we found increasing age was associated with higher hazard of mortality, evidence suggests adult patients younger than 20 years of age are at higher hazard of mortality as compared to those over the age of 20 (HR 1.37, 95% CI 1.16 – 1.60). Moderate to high certainty evidence suggested association by chance alone for cystic fibrosis vs obstructive lung disease, wedge pressure, systolic PA pressure, PaO_2/FiO_2, length of ICU stay, length of mechanical ventilation, and history of arrhythmias (table 1).

Meta-analyses of post-transplant complications

Studies commonly evaluated nine early post-operative complications. With moderate to high certainty, we noted increases in the hazard of mortality with PGD (HR 3.18, 95% CI 2.92 – 3.47), severe PGD (HR 1.66, 95% CI 1.09 – 2.51), bleeding requiring re-exploration (HR 1.22, 95% CI 1.10

– 1.35), need for ECMO (HR 1.91, 95% CI 1.79 to 2.04), need for vasopressin (HR 2.12, 95% CI 1.08 – 4.13), need for dialysis (HR 7.87, 95% CI 6.79 – 9.12), and development of cardiac complications (HR 1.25, 95% CI 1.13 – 1.39). With moderate certainty, we excluded an association between acute rejection and 1-year mortality (table 1).

Predictors addressed in a single study

We developed a repository of predictors evaluated in only one study that included factors related to the donor, transplant process, recipient comorbidities, recipient traits, recipient physiology, recipient psycho-social status, and post-transplant complications. The supplementary material presents all predictors and their associations.

DISCUSSION

Principle findings

In this review we identified, with moderate to high certainty, donor sex, type of transplant, CMV mismatch, CPB use, recipient age, recipient hypertension, recipient BMI, obstructive aetiology, coronary artery disease, post-transplant primary graft dysfunction, re-exploration for bleeding, need for ECMO, need for vasopressin, dialysis requirement, and cardiac complications to be associated with the risk of 1-year mortality post lung transplantation. Also with moderate to high certainty, we excluded associations with 1-year mortality for donor age and smoking, ischemic time, sex mismatch, HLA-A locus 2 mismatch, ABO incompatibility, recipient diabetes, cystic fibrosis, pre-transplant wedge pressure, pre-transplant systolic PA pressure, pre-transplant PaO2/FiO2 ratio, pre-transplant time on mechanical ventilation, pre-transplant arrhythmia, and post-transplant acute rejection.

Strengths and limitations

This is the first systematic review and meta-analysis addressing the adjusted associations for all predictors of 1-year mortality post lung transplantation. Through this work, we devised a

repository of all predictors for 1-year mortality. The predictors amendable to meta-analysis provided us with the opportunity to explore the sources of inconsistencies across studies, if any, through subgroup analyses. The hazard ratios obtained from these meta-analyses represent the most precise and least biased measure of association between each prognostic factor and 1-year mortality. We evaluated the certainty of the evidence using the GRADE system and found evidence consistently of high or moderate quality.

One particular limitation of this review is the methods of statistical analysis conducted by the primary studies. The authors assessed the association between predictors and mortality using Cox regression analysis. Most failed to address whether results met the proportional hazards assumption nor did they report information that would have allowed us to make that assessment.

We included UNOS registry studies to represent all studies published from individual centers in the USA rendering this review vulnerable to limitations of the UNOS registry data. It may be that authors of single- or multi-centered observational studies have better control of the collected information compared to large registries[60]. If so, non-registry studies are more likely to ensure data quality prior to analysis of risk factors[60].

Studies varied considerably in the covariates included in their regression models (studies included 12 $\pm$ 9 covariates in their regression models). Thus, results are vulnerable to the possibility that the impact of a particular prognostic factor might differ depending on which variables were included in a particular model.

In the context of identifying factors that increase the risk of 1-year mortality, the studies in this review have a fundamental limitation: potential candidates for transplant and donors may be rejected because of factors that were not included. The reasons for not utilizing the donor or

recommending transplant in such individuals may be the most powerful prognostic factors. These may include, for example: active infections, a combination of older age with a constellation of other comorbidities such as obesity, cardiovascular disease, malignancies, or an assessment of overall frailty[61].

Relation to other work

The results of this review complement the annual registry analysis conducted by the International Society of Heart and Lung Transplantation[1]. In addition to the risk factors identified by the ISHLT registry, this systematic review provides the magnitude of associations for novel prognostic factors not captured by the registry and the quality of the evidence. For example, this review provides the magnitude of association for donor sex, CMV mismatches, CPB use, recipient history of CAD, recipient hypertension, recipient underweight BMI, recipient age less than 20, recipient cystic fibrosis etiology, wedge pressure, systolic PAP, PaO_2/FiO_2, length of ICU stay, length of pre-transplant mechanical ventilation, history of arrhythmias, post-transplant PGD, bleeding requiring re-exploration, need for ECMO, need for vasopressin, need for dialysis, cardiac complications, and acute rejection. Our findings regarding the associations for these additional risk factors, in conjunction with those reported by the ISHLT, provide health care professionals and patients with a more complete picture of prognosis when evaluating a potential donor and recipient match.

We found similar and overlapping association for prognostic factors also identified by the ISHLT registry. For example, the ISHLT registry reports a HR of 0.87 (95% CI 0.81 – 0.93) for BLTx compared to SLTx[1]. For the same risk factor, we observed a HR of 0.81 (95% CI 0.75 – 0.87). For all prognostic factors, we translate the relative effects into absolute risks, which is useful for understanding the impact of each factor at the point of care (absolute risk increase or reduction).

Such absolute risk measures become important in identifying prognostic factors that make a large impact on the risk of 1-year mortality, compared to those that make a trivial impact[11].

The ISHLT registry observed a statistically significant association between ischemic time and 1-year mortality. The registry conducted their analysis using cubic splines, which allows flexibility in the linearity assumption of regression models[1]. The studies included in our meta-analysis did not utilize cubic splines. Cubic splines are useful when there is a certain threshold for a continuous prognostic factor, at which the risk significantly increases. Based on the registry analysis, this point seems to be an ischemic time of 5 hours[1]. In our list of unique predictors, we identified two studies that evaluated ischemic time as a binary variable (threshold of 6 hours and 8 hours). The study evaluating a threshold of 6 hours reported a HR 1.20 (95% CI 1.02 – 1.42)[62] and the one evaluating a threshold of 8 hours reported a HR of 1.29 (95% CI 1.19 – 1.39)[51].

Unlike the ISHLT registry analysis, we did not observe an association beyond chance for donor age. Our pooled HR and 95% CI was informed by 2 studies including 21,078 patients[24 54]. The median donor age included in our review was 34 years. The ISHLT registry analysis, once again using cubic splines, observed a statistically significant association between donor age and 1-year mortality when the donor age was greater than 40 years. The discrepancy between our results and ISHLT may be due to the limited range of donor age in the studies of this review, and the lack of consideration for the non-linear association between donor age and 1-year mortality (as highlighted by the cubic spline transformation applied by the ISHLT)[1]. The implication is that donor ages outside of the ranges of these studies might be worthy of consideration for lung transplantation. Use of ex-vivo lung perfusion to assess suitability of organs further supports the desirability of not discarding donors based on age alone[63 64].

Implication for guidelines

The ISHLT provides guidance on the selection of candidates for lung transplantation. In general lung transplantation is considered for those with >50% risk of mortality from lung disease within 2-years, >80% likelihood of surviving 90 days post-transplant, and >80% likelihood of 5-year survival from a general medical perspective provided that there is adequate graft function[61]. The current review suggests that numerous recipient and donor characteristics increase the risk of mortality post lung transplantation. Careful examination of absolute risk increases of each prognostic factor (table 1) suggests that although there may be a statistically significant association beyond chance, each factor alone may not appreciably diminish the magnitude of benefit attained from transplantation. More specifically, none of the identified and meta-analyzed risk factors decrease the likelihood of 90-day or 1-year survival below the 80% threshold. This necessitates the need for risk prediction models to guide clinicians in selection of candidates whose risk for mortality may be higher than the aforementioned thresholds. Risk associations generated from this review may inspire or provide the foundational information necessary for development of a risk prediction model.

CONCLUSION

Our systematic review and meta-analysis identified 20 prognostic factors for which we have moderate or high certainty in their magnitude of association with 1-year mortality post lung transplantation. These factors include donor sex, bilateral or single lung transplant, CMV mismatch, CPB use, recipient age, recipient hypertension, recipient underweight BMI, recipient etiology of lung disease, PaO_2/FiO_2 pressure, coronary artery disease, development of PGD post-transplant, severe bleeding, need for ECMO, need for vasopressin, need for dialysis, and any cardiac complication post-transplant. The identified prognostic factors may be used in the

development of future risk prediction models, that may in turn guide the judgment clinicians need to make on the highest risk recipients and donors.

ACKNOWLEDGEMENTS

The authors acknowledge Ms. Ani Orchania-Cheff for developing the search strategy for this review.

DISCLOSURES

None of the authors have any conflicts of interest to disclose.

REFERENCE

1. Chambers DC, Cherikh WS, Harhay MO, et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-sixth adult lung and heart-lung transplantation Report-2019; Focus theme: Donor and recipient size match. *J Heart Lung Transplant* 2019;38(10):1042-55. doi: 10.1016/j.healun.2019.08.001 [published Online First: 2019/09/25]
2. Krumholz HM. Registries and selection bias: the need for accountability. *Circ Cardiovasc Qual Outcomes* 2009;2(6):517-8. doi: 10.1161/CIRCOUTCOMES.109.916601 [published Online First: 2009/12/25]
3. Williams WG, McCrindle BW. Practical experience with databases for congenital heart disease: a registry versus an academic database. *Semin Thorac Cardiovasc Surg Pediatr Card Surg Annu* 2002;5:132-42. doi: 10.1053/pcsu.2002.31485 [published Online First: 2002/05/08]
4. Borro JM, Delgado M, Coll E, et al. Single-lung transplantation in emphysema: Retrospective study analyzing survival and waiting list mortality. *World J Transplant* 2016;6(2):347-55. doi: 10.5500/wjt.v6.i2.347 [published Online First: 2016/07/01]
5. Jacques F, El-Hamamsy I, Fortier A, et al. Acute renal failure following lung transplantation: risk factors, mortality, and long-term consequences. *Eur J Cardiothorac Surg* 2012;41(1):193-9. doi: 10.1016/j.ejcts.2011.04.034 [published Online First: 2011/06/15]
6. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 2013;10(2):e1001380. doi: 10.1371/journal.pmed.1001380 [published Online First: 2013/02/09]
7. Foroutan F, Guyatt G, Friesen E, et al. Predictors of 1-year mortality in adult lung transplant recipients: a systematic review and meta-analysis. *Syst Rev* 2019;8(1):131. doi: 10.1186/s13643-019-1049-x [published Online First: 2019/06/05]
8. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6. doi: 10.7326/0003-4819-158-4-201302190-00009 [published Online First: 2013/02/20]
9. Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *BMJ* 2014;348:f7450. doi: 10.1136/bmj.f7450 [published Online First: 2014/01/28]
10. Tierney JF, Stewart LA, Ghersi D, et al. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;8(1):16. doi: 10.1186/1745-6215-8-16
11. Foroutan F, Iorio A, Thabane L, et al. Calculation of absolute risk for important outcomes in patients with and without a prognostic factor of interest. *J Clin Epidemiol* 2019 doi: 10.1016/j.jclinepi.2019.08.012 [published Online First: 2019/10/08]
12. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ : British Medical Journal* 2015;350:h870. doi: 10.1136/bmj.h870
13. Harris RJ, Deeks JJ, Altman DG, et al. Metan: Fixed- and Random-Effects Meta-Analysis. *The Stata Journal* 2008;8(1):3-28. doi: 10.1177/1536867X0800800102
14. Foroutan F, Guyatt G, Victoia Z, et al. Use of GRADE for assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks. *Journal of Clinical Epidemiology* 2020

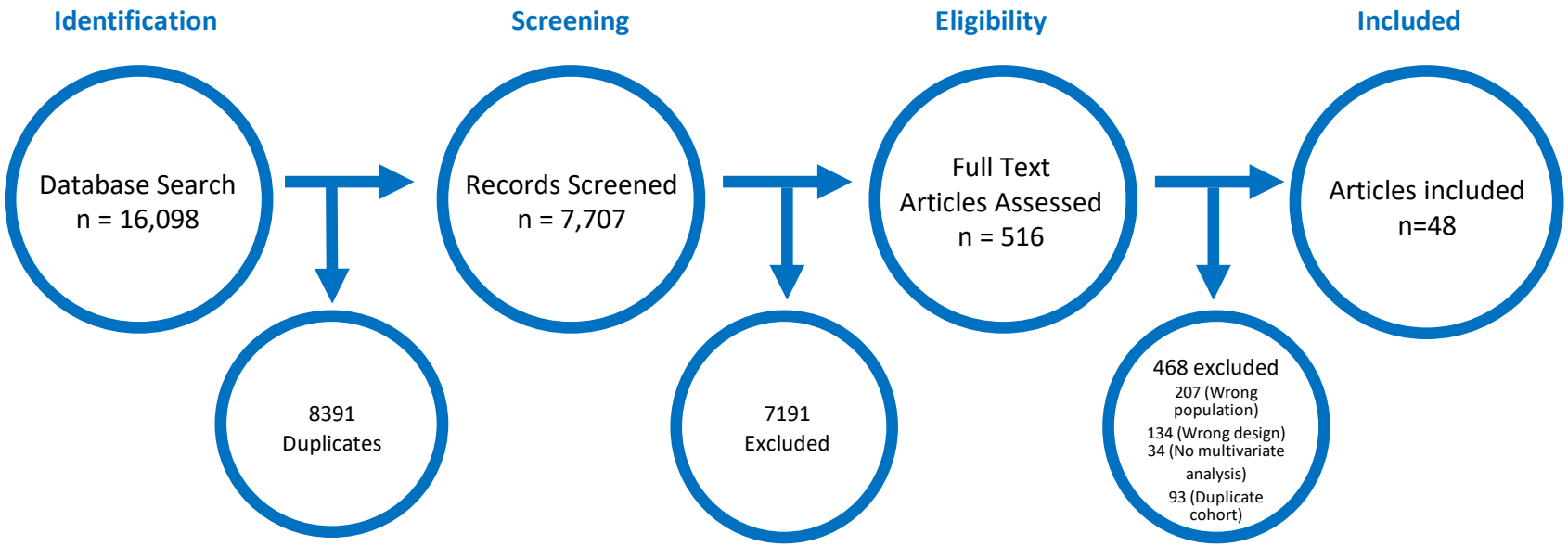
15. Allen JG, Arnaoutakis GJ, Orens JB, et al. Insurance status is an independent predictor of long-term survival after lung transplantation in the United States. *J Heart Lung Transplant* 2011;30(1):45-53. doi: 10.1016/j.healun.2010.07.003 [published Online First: 2010/09/28]
16. Allen JG, Arnaoutakis GJ, Weiss ES, et al. The impact of recipient body mass index on survival after lung transplantation. *J Heart Lung Transplant* 2010;29(9):1026-33. doi: 10.1016/j.healun.2010.05.005 [published Online First: 2010/06/19]
17. Allyn PR, Duffy EL, Humphries RM, et al. Graft Loss and CLAD-Onset Is Hastened by Viral Pneumonia After Lung Transplantation. *Transplantation* 2016;100(11):2424-31. doi: 10.1097/TP.0000000000001346 [published Online First: 2016/10/25]
18. Arnaoutakis GJ, George TJ, Robinson CW, et al. Severe acute kidney injury according to the RIFLE (risk, injury, failure, loss, end stage) criteria affects mortality in lung transplantation. *J Heart Lung Transplant* 2011;30(10):1161-8. doi: 10.1016/j.healun.2011.04.013 [published Online First: 2011/05/31]
19. Awori Hayanga JW, Aboagye JK, Shigemura N, et al. Airway complications after lung transplantation: Contemporary survival and outcomes. *J Heart Lung Transplant* 2016;35(10):1206-11. doi: 10.1016/j.healun.2016.04.019 [published Online First: 2016/10/25]
20. Bonser RS, Taylor R, Collett D, et al. Effect of donor smoking on survival after lung transplantation: a cohort study of a prospective registry. *Lancet* 2012;380(9843):747-55. doi: 10.1016/S0140-6736(12)60160-3 [published Online First: 2012/06/01]
21. Cassivi SD, Meyers BF, Battafarano RJ, et al. Thirteen-year experience in lung transplantation for emphysema. *Ann Thorac Surg* 2002;74(5):1663-9; discussion 69-70. doi: 10.1016/s0003-4975(02)04064-x [published Online First: 2002/11/21]
22. Castleberry AW, Martin JT, Osho AA, et al. Coronary revascularization in lung transplant recipients with concomitant coronary artery disease. *Am J Transplant* 2013;13(11):2978-88. doi: 10.1111/ajt.12435 [published Online First: 2013/10/10]
23. Ceriana P, Klersy C, Veronesi R, et al. Influence of underlying lung disease on early postoperative course after single lung transplantation. *J Cardiovasc Surg (Torino)* 2002;43(5):715-22. [published Online First: 2002/10/19]
24. Ceron Navarro J, de Aguiar Quevedo K, Mancheno Franch N, et al. [Complications after lung transplantation in chronic obstructive pulmonary disease]. *Med Clin (Barc)* 2013;140(9):385-9. doi: 10.1016/j.medcli.2012.07.028 [published Online First: 2013/03/07]
25. Chaikriangkrai K, Jhun HY, Graviss EA, et al. Overweight-mortality paradox and impact of six-minute walk distance in lung transplantation. *Ann Thorac Med* 2015;10(3):169-75. doi: 10.4103/1817-1737.160835 [published Online First: 2015/08/01]
26. Chaikriangkrai K, Jyothula S, Jhun HY, et al. Incidence, Risk Factors, Prognosis, and Electrophysiological Mechanisms of Atrial Arrhythmias after Lung Transplantation. *JACC Clin Electrophysiol* 2015;1(4):296-305. doi: 10.1016/j.jacep.2015.05.009 [published Online First: 2015/11/12]
27. Chandrashekaran S, Keller CA, Kremers WK, et al. Weight loss prior to lung transplantation is associated with improved survival. *J Heart Lung Transplant* 2015;34(5):651-7. doi: 10.1016/j.healun.2014.11.018 [published Online First: 2015/01/13]

28. Christie JD, Kotloff RM, Ahya VN, et al. The effect of primary graft dysfunction on survival after lung transplantation. *Am J Respir Crit Care Med* 2005;171(11):1312-6. doi: 10.1164/rccm.200409-1243OC [published Online First: 2005/03/15]
29. D'Angelo AM, Chan EG, Hayanga JW, et al. Atrial arrhythmias after lung transplantation: Incidence and risk factors in 652 lung transplant recipients. *J Thorac Cardiovasc Surg* 2016;152(3):901-9. doi: 10.1016/j.jtcvs.2016.04.070 [published Online First: 2016/05/29]
30. Demir A, Coosemans W, Decaluwe H, et al. Donor-recipient matching in lung transplantation: which variables are important?dagger. *Eur J Cardiothorac Surg* 2015;47(6):974-83. doi: 10.1093/ejcts/ezu340 [published Online First: 2014/09/05]
31. DerHovanessian A, Todd JL, Zhang A, et al. Validation and Refinement of Chronic Lung Allograft Dysfunction Phenotypes in Bilateral and Single Lung Recipients. *Ann Am Thorac Soc* 2016;13(5):627-35. doi: 10.1513/AnnalsATS.201510-719OC [published Online First: 2016/05/06]
32. George TJ, Arnaoutakis GJ, Beaty CA, et al. Acute kidney injury increases mortality after lung transplantation. *Ann Thorac Surg* 2012;94(1):185-92. doi: 10.1016/j.athoracsur.2011.11.032 [published Online First: 2012/02/14]
33. Hayanga JA, Lira A, Vlahu T, et al. Lung Transplantation in Patients with High Lung Allocation Scores in the US: Evidence for the Need to Evaluate Score Specific Outcomes. *J Transplant* 2015;2015:836751. doi: 10.1155/2015/836751 [published Online First: 2016/01/23]
34. Hayes D, Jr., Black SM, Tobias JD, et al. Influence of donor and recipient age in lung transplantation. *J Heart Lung Transplant* 2015;34(1):43-9. doi: 10.1016/j.healun.2014.08.017 [published Online First: 2014/10/11]
35. Hayes D, Jr., Tobias JD, Tumin D. Center Volume and Extracorporeal Membrane Oxygenation Support at Lung Transplantation in the Lung Allocation Score Era. *Am J Respir Crit Care Med* 2016;194(3):317-26. doi: 10.1164/rccm.201511-2222OC [published Online First: 2016/02/04]
36. Hayes D, Jr., Whitson BA, Ghadiali SN, et al. Influence of HLA Mismatching on Survival in Lung Transplantation. *Lung* 2015;193(5):789-97. doi: 10.1007/s00408-015-9768-9 [published Online First: 2015/07/30]
37. Henri C, Giraldeau G, Dorais M, et al. Atrial fibrillation after pulmonary transplantation: incidence, impact on mortality, treatment effectiveness, and risk factors. *Circ Arrhythm Electrophysiol* 2012;5(1):61-7. doi: 10.1161/CIRCEP.111.964569 [published Online First: 2011/12/14]
38. Huppmann P, Neurohr C, Leuschner S, et al. The Munich-LTX-Score: predictor for survival after lung transplantation. *Clin Transplant* 2012;26(1):173-83. doi: 10.1111/j.1399-0012.2011.01573.x [published Online First: 2011/12/23]
39. Inci I, Schuurmans M, Ehrsam J, et al. Lung transplantation for emphysema: impact of age on short- and long-term survival. *Eur J Cardiothorac Surg* 2015;48(6):906-9. doi: 10.1093/ejcts/ezu550 [published Online First: 2015/01/21]
40. Julliard WA, Meyer KC, De Oliveira NC, et al. The presence or severity of pulmonary hypertension does not affect outcomes for single-lung transplantation. *Thorax* 2016;71(5):478-80. doi: 10.1136/thoraxjnl-2015-207354 [published Online First: 2015/12/02]

41. Loor G, Brown R, Kelly RF, et al. Gender differences in long-term survival post-transplant: A single-institution analysis in the lung allocation score era. *Clin Transplant* 2017;31(3) doi: 10.1111/ctr.12889 [published Online First: 2016/12/19]
42. Mason DP, Batizy LH, Wu J, et al. Matching donor to recipient in lung transplantation: How much does size matter? *J Thorac Cardiovasc Surg* 2009;137(5):1234-40 e1. doi: 10.1016/j.jtcvs.2008.10.024 [published Online First: 2009/04/22]
43. Minambres E, Zurbano F, Naranjo S, et al. [Lung transplantation using donors 55 years and older]. *Med Intensiva* 2011;35(7):403-9. doi: 10.1016/j.medin.2010.12.006 [published Online First: 2011/02/24]
44. Minambres E, Zurbano F, Naranjo S, et al. [Mortality analysis of patients undergoing lung transplantation for emphysema]. *Arch Bronconeumol* 2009;45(7):335-40. doi: 10.1016/j.arbres.2009.01.011 [published Online First: 2009/05/27]
45. Mollberg NM, Farjah F, Howell E, et al. Impact of primary caregivers on long-term outcomes after lung transplantation. *J Heart Lung Transplant* 2015;34(1):59-64. doi: 10.1016/j.healun.2014.09.022 [published Online First: 2014/12/03]
46. Neurohr C, Huppmann P, Thum D, et al. Potential functional and survival benefit of double over single lung transplantation for selected patients with idiopathic pulmonary fibrosis. *Transpl Int* 2010;23(9):887-96. doi: 10.1111/j.1432-2277.2010.01071.x [published Online First: 2010/03/17]
47. Newton CA, Kozlitina J, Lines JR, et al. Telomere length in patients with pulmonary fibrosis associated with chronic lung allograft dysfunction and post-lung transplantation survival. *J Heart Lung Transplant* 2017;36(8):845-53. doi: 10.1016/j.healun.2017.02.005 [published Online First: 2017/03/07]
48. Plantier L, Skhiri N, Biondi G, et al. Impact of previous cardiovascular disease on the outcome of lung transplantation. *J Heart Lung Transplant* 2010;29(11):1270-6. doi: 10.1016/j.healun.2010.05.010 [published Online First: 2010/06/29]
49. Schaffer JM, Singh SK, Reitz BA, et al. Single- vs double-lung transplantation in patients with chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis since the implementation of lung allocation based on medical need. *JAMA* 2015;313(9):936-48. doi: 10.1001/jama.2015.1175 [published Online First: 2015/03/04]
50. Shigemura N, Orhan Y, Bhama JK, et al. Delayed chest closure after lung transplantation: techniques, outcomes, and strategies. *J Heart Lung Transplant* 2014;33(7):741-8. doi: 10.1016/j.healun.2014.03.003 [published Online First: 2014/04/15]
51. Shigemura N, Sclabassi RJ, Bhama JK, et al. Early major neurologic complications after lung transplantation: incidence, risk factors, and outcome. *Transplantation* 2013;95(6):866-71. doi: 10.1097/TP.0b013e318280b359 [published Online First: 2013/01/29]
52. Shigemura N, Toyoda Y, Bhama JK, et al. Donor smoking history and age in lung transplantation: a revisit. *Transplantation* 2013;95(3):513-8. doi: 10.1097/TP.0b013e3182751f1f [published Online First: 2012/12/04]
53. Shino MY, Weigt SS, Li N, et al. CXCR3 ligands are associated with the continuum of diffuse alveolar damage to chronic lung allograft dysfunction. *Am J Respir Crit Care Med* 2013;188(9):1117-25. doi: 10.1164/rccm.201305-0861OC [published Online First: 2013/09/26]

54. Taimeh Z, Hertz MI, Shumway S, et al. Lung transplantation for pulmonary sarcoidosis. Twenty-five years of experience in the USA. *Thorax* 2016;71(4):378-9. doi: 10.1136/thoraxjnl-2015-207497 [published Online First: 2016/01/20]
55. Vadnerkar A, Toyoda Y, Crespo M, et al. Age-specific complications among lung transplant recipients 60 years and older. *J Heart Lung Transplant* 2011;30(3):273-81. doi: 10.1016/j.healun.2010.08.032 [published Online First: 2010/12/15]
56. Wehbe E, Brock R, Budev M, et al. Short-term and long-term outcomes of acute kidney injury after lung transplantation. *J Heart Lung Transplant* 2012;31(3):244-51. doi: 10.1016/j.healun.2011.08.016 [published Online First: 2011/10/15]
57. Whitson BA, Hertz MI, Kelly RF, et al. Use of the donor lung after asphyxiation or drowning: effect on lung transplant recipients. *Ann Thorac Surg* 2014;98(4):1145-51. doi: 10.1016/j.athoracsur.2014.05.065 [published Online First: 2014/08/20]
58. Xue J, Wang L, Chen CM, et al. Acute kidney injury influences mortality in lung transplantation. *Ren Fail* 2014;36(4):541-5. doi: 10.3109/0886022X.2013.876350 [published Online First: 2014/04/20]
59. Zalunardo MP, Thalmann C, Seifert B, et al. Impact of preoperative right-ventricular function and platelet transfusion on outcome after lung transplantation. *Eur J Cardiothorac Surg* 2011;39(4):538-42. doi: 10.1016/j.ejcts.2010.07.040 [published Online First: 2010/12/15]
60. Ferreira-Gonzalez I, Marsal JR, Mitjavila F, et al. Patient registries of acute coronary syndrome: assessing or biasing the clinical real world data? *Circ Cardiovasc Qual Outcomes* 2009;2(6):540-7. doi: 10.1161/CIRCOUTCOMES.108.844399 [published Online First: 2009/12/25]
61. Weill D, Benden C, Corris PA, et al. A consensus document for the selection of lung transplant candidates: 2014--an update from the Pulmonary Transplantation Council of the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant* 2015;34(1):1-15. doi: 10.1016/j.healun.2014.06.014 [published Online First: 2014/08/03]
62. Shigemura N, Bhama J, Gries CJ, et al. Lung transplantation in patients with prior cardiothoracic surgical procedures. *Am J Transplant* 2012;12(5):1249-55. doi: 10.1111/j.1600-6143.2011.03946.x [published Online First: 2012/02/04]
63. Valenza F, Rosso L, Coppola S, et al. Ex vivo lung perfusion to improve donor lung function and increase the number of organs available for transplantation. *Transpl Int* 2014;27(6):553-61. doi: 10.1111/tri.12295 [published Online First: 2014/03/19]
64. Possoz J, Neyrinck A, Van Raemdonck D. Ex vivo lung perfusion prior to transplantation: an overview of current clinical practice worldwide. *J Thorac Dis* 2019;11(4):1635-50. doi: 10.21037/jtd.2019.04.33 [published Online First: 2019/06/11]

Figure 1 – PRISAM flow diagram for selection of eligible articles



Wrong design: randomized controlled trials or studies, case-control studies, case-series, or editorials/commentaries.
Wrong population: Excluding patients with early mortality (any study that excluded patients with survival less than 1-year).
Recipients of multi-organ transplants or undergoing re-transplantation.
Duplicate cohorts: two studies with the same cohort reporting on the same predictor and exact same outcome.

Figure 2 – All predictors of 1-year mortality amenable to meta-analysis

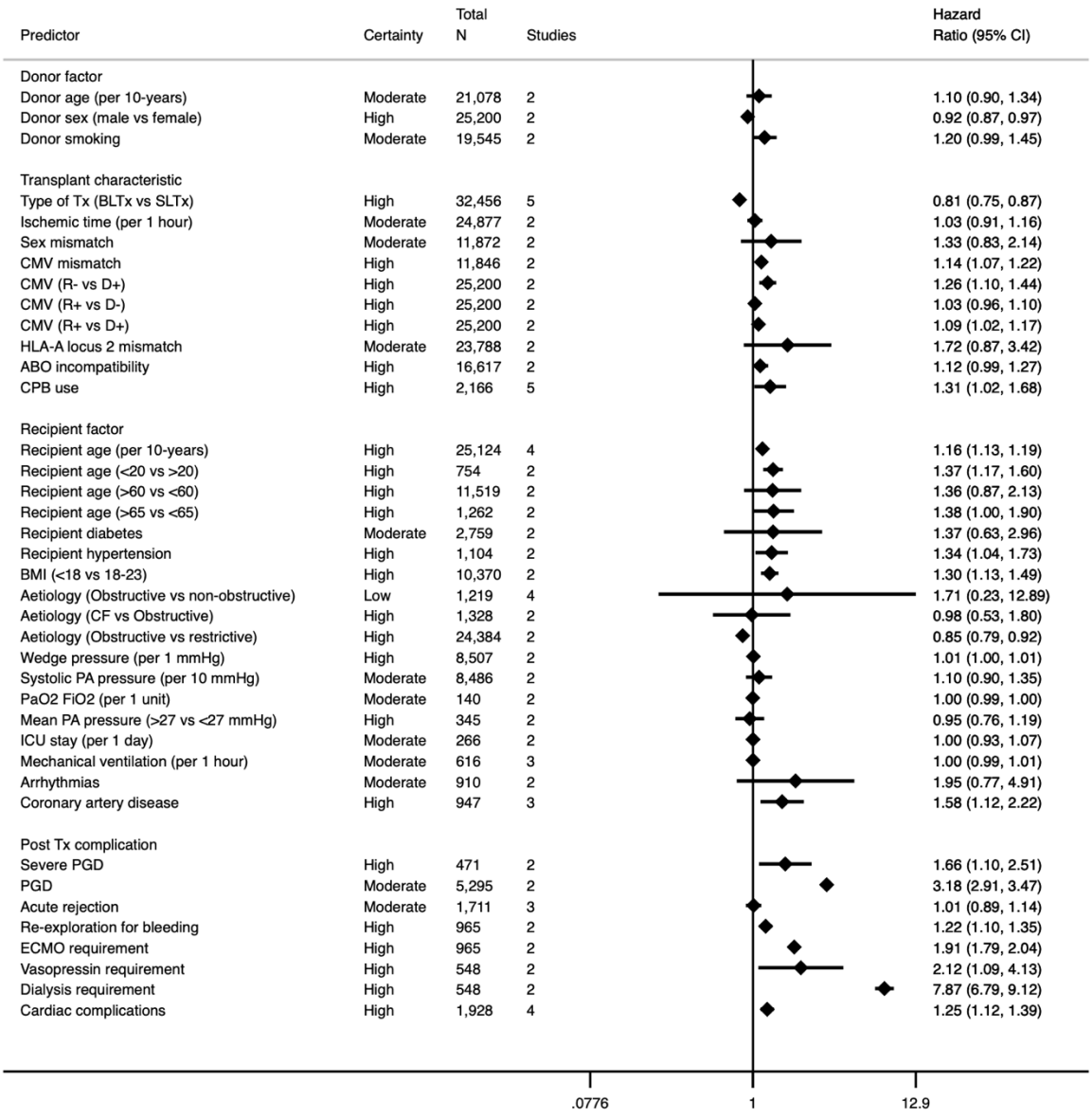


Table 1 – Summary of Findings Tables

Predictor	Study Results and measurements	Prevalence	Absolute effect estimates Baseline	Absolute effect estimates With predictor	Certainty in effect estimates (Quality of evidence)	Plain text summary
Donor Factors						
Donor Age (Per 10 years)	Hazard Ratio 1.10 (CI 95% 0.90 - 1.34) Based on data from 21,078 patients in 2 studies	N/A	158 per 1000	172 per 1000	Moderate Due to serious risk of bias	probably has little or no effect on 1-year mortality
			Difference: 14 more per 1000 (CI 95% 15 fewer - 48 more)			
Donor Sex (Male vs Female)	Hazard Ratio 0.92 (CI 95% 0.88 - 0.97) Based on data from 25,200 patients in 2 studies	60%	165 per 1000	153 per 1000	High	slightly decreases the risk of 1-year mortality
			Difference: 12 fewer per 1000 (CI 95% 19 fewer - 5 fewer)			
Donor smoking (Yes vs No)	Hazard Ratio 1.20 (CI 95% 1.00 - 1.45) Based on data from 19,545 patients in 2 studies	31%	150 per 1000	177 per 1000	Low Due to serious inconsistency and imprecision	probably increases the risk of 1-year mortality
			Difference: 27 more per 1000 (CI 95% 0 more - 56 more)			
Transplant Characteristics						
Type of Tx (BLTx vs SLTx)	Hazard Ratio 0.81 (CI 95% 0.75 - 0.87) Based on data from 32,456 patients in 5 studies	57%	176 per 1000	145 per 1000	High	slightly decreases the risk of 1-year mortality
			Difference: 31 fewer per 1000 (CI 95% 42 fewer - 21 fewer)			
Ischemic Time (Per hour)	Hazard Ratio 1.03 (CI 95% 0.91 - 1.16) Based on data from 24,877 patients in 2 studies	N/A	158 per 1000	162 per 1000	Moderate Due to serious inconsistency	probably has little or no effect on 1-year mortality
			Difference: 4 more per 1000 (CI 95% 13 fewer - 23 more)			
Sex mismatch (Yes vs No)	Hazard Ratio 1.33 (CI 95% 0.82 - 2.14) Based on data from 11,872 patients in 2 studies	31%	145 per 1000	187 per 1000	Low Due to serious inconsistency and imprecision	probably has little or no effect on 1-year mortality
			Difference: 42 more per 1000 (CI 95% 28 fewer - 116 more)			
CMV Status (Mismatch vs Match)	Hazard Ratio 1.14 (CI 95% 1.07 - 1.22) Based on data from 11,846 patients in 2 studies	42%	150 per 1000	169 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 19 more per 1000 (CI 95% 10 more - 29 more)			
CMV Status (R -, D +)	Hazard Ratio 1.26 (CI 95% 1.11 - 1.44) Based on data from 25,200 patients in 2 studies	16%	152 per 1000	188 per 1000	Moderate Due to serious imprecision	slightly increases the risk of 1-year mortality
			Difference: 36 more per 1000 (CI 95% 15 more - 58 more)			
CMV Status (R +, D -)	Hazard Ratio 1.03 (CI 95% 0.96 - 1.11) Based on data from 25,200 patients in 2 studies	22%	157 per 1000	161 per 1000	High	has little or no effect on 1-year mortality
			Difference: 4 more per 1000 (CI 95% 6 fewer - 15 more)			
CMV Status (R +, D +)	Hazard Ratio 1.09 (CI 95% 1.02 - 1.17) Based on data from 25,200 patients in 2 studies	22%	155 per 1000	168 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 13 more per 1000 (CI 95% 3 more - 24 more)			
HLA -A locus 2 (Mismatch vs Match)	Hazard Ratio 1.72 (CI 95% 0.87 - 3.42) Based on data from 23,788 patients in 2 studies	46%	123 per 1000	199 per 1000	Low Due to serious risk of bias and imprecision	probably has little or no effect on 1-year mortality
			Difference: 76 more per 1000 (CI 95% 20 fewer - 157 more)			
ABO Compatibility (No vs Yes)	Hazard Ratio 1.12 (CI 95% 0.98 - 1.27) Based on data from 16,617 patients in 2 studies	17%	155 per 1000	172 per 1000	High	has little or no effect on 1-year mortality
			Difference: 17 more per 1000 (CI 95% 2 fewer - 36 more)			
CPB use (Yes vs No)	Hazard Ratio 1.31 (CI 95% 1.03 - 1.68) Based on data from 2,166 patients in 5 studies	22%	149 per 1000	190 per 1000	Moderate Due to serious imprecision	slightly increases the risk of 1-year mortality
			Difference: 41 more per 1000 (CI 95% 4 more - 82 more)			

Table 1 Continued – Summary Findings Table

Predictor	Study Results and measurements	Prevalence	Absolute effect estimates		Certainty in effect estimates (Quality of evidence)	Plain text summary
			Baseline	With predictor		
Recipient Factors						
Recipient Age (per 10 years)	Hazard Ratio 1.16 (CI 95% 1.13 - 1.19) Based on data from 25,124 patients in 4 studies	N/A	158 per 1000	181 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 23 more per 1000 (CI 95% 19 fewer - 27 more)			
Recipient Age (<20 vs >20)	Hazard Ratio 1.37 (CI 95% 1.16 - 1.60) Based on data from 754 patients in 2 studies	4%	156 per 1000	207 per 1000	Moderate Due to serious imprecision	slightly increases the risk of 1-year mortality
			Difference: 51 more per 1000 (CI 95% 23 more - 81 more)			
Recipient Age (>60 vs <60)	Hazard Ratio 1.36 (CI 95% 0.87 - 2.13) Based on data from 11,519 patients in 2 studies	32%	143 per 1000	189 per 1000	Moderate Due to serious imprecision	has little or no effect on 1-year mortality
			Difference: 46 more per 1000 (CI 95% 19 fewer - 115 more)			
Recipient Age (>65 vs <65)	Hazard Ratio 1.38 (CI 95% 1.00 - 1.90) Based on data from 1,262 patients in 2 studies	17%	149 per 1000	200 per 1000	Moderate Due to serious imprecision	probably has little or no effect on 1-year mortality
			Difference: 51 more per 1000 (CI 95% 0 more - 107 more)			
Recipient sex (Female vs Male)	Hazard Ratio 1.03 (CI 95% 0.88 - 1.20) Based on data from 24,613 patients in 4 studies	56%	155 per 1000	160 per 1000	High	has little or no effect on 1-year mortality
			Difference: 5 more per 1000 (CI 95% 20 fewer - 29 more)			
Recipient diabetes (Yes vs No)	Hazard Ratio 1.37 (CI 95% 0.64 - 2.96) Based on data from 2,759 patients in 2 studies	15%	151 per 1000	200 per 1000	Low Due to serious risk of bias and imprecision	probably has little or no effect on 1-year mortality
			Difference: 29 more per 1000 (CI 95% 56 fewer - 197 more)			
Recipient hypertension (Yes vs No)	Hazard Ratio 1.34 (CI 95% 1.04 - 1.73) Based on data from 1,104 patients in 2 studies	17%	150 per 1000	196 per 1000	Moderate Due to serious imprecision	slightly increases the risk of 1-year mortality
			Difference: 46 more per 1000 (CI 95% 6 more - 90 more)			
BMI (<18 vs 18 - 23)	Hazard Ratio 1.30 (CI 95% 1.14 - 1.49) Based on data from 10,370 patients in 2 studies	15%	152 per 1000	193 per 1000	Moderate Due to serious imprecision	slightly increases the risk of 1-year mortality
			Difference: 41 more per 1000 (CI 95% 20 more - 64 more)			
Aetiology (Obstructive vs Non-obstructive)	Hazard Ratio 1.71 (CI 95% 0.23 - 12.89) Based on data from 1,219 patients in 4 studies	37%	129 per 1000	208 per 1000	Low Due to serious risk of bias and imprecision	may have little or no effect on the risk of 1-year mortality
			Difference: 79 more per 1000 (CI 95% 165 fewer - 284 more)			
Aetiology (CF vs Obstructive)	Hazard Ratio 0.98 (CI 95% 0.51 - 1.88) Based on data from 1,328 patients in 2 studies	14%	158 per 1000	155 per 1000	Moderate Due to serious imprecision	has little or no effect on 1-year mortality
			Difference: 3 fewer per 1000 (CI 95% 79 fewer - 107 more)			
Aetiology (Obstructive vs Restrictive)	Hazard Ratio 0.85 (CI 95% 0.78 - 0.92) Based on data from 24,384 patients in 2 studies	37%	167 per 1000	143 per 1000	High	slightly decreases the risk of 1-year mortality
			Difference: 24 fewer per 1000 (CI 95% 35 fewer - 12 fewer)			
Wedge pressure (per 1 mmHg)	Hazard Ratio 1.01 (CI 95% 1.00 - 1.01) Based on data from 8,507 patients in 2 studies	N/A	158 per 1000	159 per 1000	High	has little or no effect on 1-year mortality
			Difference: 1 more per 1000 (CI 95% 0 more - 1 more)			
Systolic PA pressure (per 10 mmHg)	Hazard Ratio 1.10 (CI 95% 0.89 - 1.35) Based on data from 8,486 patients in 2 studies	N/A	158 per 1000	172 per 1000	Moderate Due to serious risk of bias	probably has little or no effect on 1-year mortality
			Difference: 14 more per 1000 (CI 95% 16 fewer - 49 more)			
PaO_2 FiO_2 (per 1 unit)	Hazard Ratio 1.00 (CI 95% 0.99 - 1.00) Based on data from 140 patients in 2 studies	N/A	158 per 1000	158 per 1000	Moderate Due to serious risk of bias	probably has little or no effect on 1-year mortality
			Difference: 0 more per 1000 (CI 95% 1 fewer - 0 more)			
mean PA pressure (>27 vs <27 mmHg)	Hazard Ratio 0.95 (CI 95% 0.76 - 1.19) Based on data from 345 patients in 2 studies	29%	160 per 1000	153 per 1000	High	has little or no effect on 1-year mortality
			Difference: 7 fewer per 1000 (CI 95% 38 fewer - 25 more)			
ICU stay (per 1 day)	Hazard Ratio 1.00 (CI 95% 0.94 - 1.07) Based on data from 266 patients in 2 studies	N/A	158 per 1000	158 per 1000	Moderate Due to serious risk of bias	probably has little or no effect on 1-year mortality
			Difference: 0 fewer per 1000 (CI 95% 9 fewer - 10 more)			
Mechanical ventilation (per 1 hour)	Hazard Ratio 1.00 (CI 95% 0.99 - 1.01) Based on data from 616 patients in 3 studies	N/A	158 per 1000	158 per 1000	Moderate Due to serious inconsistency	probably has little or no effect on 1-year mortality
			Difference: 0 fewer per 1000 (CI 95% 1 fewer - 1 more)			
Arrhythmias (Yes vs No)	Hazard Ratio 1.95 (CI 95% 0.78 - 4.91) Based on data from 910 patients in 2 studies	2%	156 per 1000	280 per 1000	Moderate Due to serious risk of bias	probably has little or no effect on 1-year mortality
			Difference: 124 more per 1000 (CI 95% 33 fewer - 393 more)			
Coronary Artery Disease (Yes vs No)	Hazard Ratio 1.58 (CI 95% 1.13 - 2.22) Based on data from 947 patients in 3 studies	16%	146 per 1000	220 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 74 more per 1000 (CI 95% 18 more - 137 more)			

Table 1 Continued – Summary Findings Table

Predictor	Study Results and measurements	Prevalence	Absolute effect estimates		Certainty in effect estimates (Quality of evidence)	Plain text summary
			Baseline	With predictor		
Post Transplant Complications						
Severe PGD (Yes vs No)	Hazard Ratio 1.66 (CI 95% 1.09 - 2.51) Based on data from 471 patients in 2 studies	11%	149 per 1000	234 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 85 more per 1000 (CI 95% 12 more - 170 more)			
PGD (Yes vs No)	Hazard Ratio 3.18 (CI 95% 2.92 - 3.47) Based on data from 5,295 patients in 2 studies	9%	137 per 1000	367 per 1000	Moderate Due to serious risk of bias	Ishcemic time per hour increase may have little or no impact on mortality
			Difference: 230 more per 1000 (CI 95% 209 more - 251 more)			
Acute Rejection (Yes vs No)	Hazard Ratio 1.01 (CI 95% 0.90 - 1.14) Based on data from 1,711 patients in 3 studies	3%	158 per 1000	159 per 1000	Moderate Due to serious inconsistency	probably has little or no effect on 1-year mortality
			Difference: 1 more per 1000 (CI 95% 14 fewer - 20 more)			
Re-exploration for bleeding (Yes vs No)	Hazard Ratio 1.22 (CI 95% 1.10 - 1.35) Based on data from 965 patients in 2 studies	5%	156 per 1000	187 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 31 more per 1000 (CI 95% 15 more - 48 more)			
ECMO requirement (Yes vs No)	Hazard Ratio 1.91 (CI 95% 1.79 - 2.04) Based on data from 965 patients in 2 studies	1%	157 per 1000	278 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 121 more per 1000 (CI 95% 106 more - 136 more)			
Vasopressin requirement (Yes vs No)	Hazard Ratio 2.12 (CI 95% 1.08 - 4.13) Based on data from 548 patients in 2 studies	84%	88 per 1000	171 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 83 more per 1000 (CI 95% 11 more - 124 more)			
Dialysis requirement (Yes vs No)	Hazard Ratio 7.87 (CI 95% 6.79 - 9.12) Based on data from 548 patients in 2 studies	5%	133 per 1000	635 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 502 more per 1000 (CI 95% 461 more - 541 more)			
Cardiac complications (Yes vs No)	Hazard Ratio 1.25 (CI 95% 1.13 - 1.39) Based on data from 1,928 patients in 4 studies	14%	153 per 1000	188 per 1000	High	slightly increases the risk of 1-year mortality
			Difference: 35 more per 1000 (CI 95% 19 more - 52 more)			
Acute Kidney Injury (Yes vs No)	Hazard Ratio 1.31 (CI 95% 0.85 - 2.28) Based on data from 851 patients in 3 studies	49%	139 per 1000	178 per 1000	Low Due to serious risk of bias and inconsistency	slightly increases the risk of 1-year mortality
			Difference: 39 more per 1000 (CI 95% 24 fewer - 110 more)			

APPENDIX A – Characteristics and demographics of included studies

First Author	Year	Multi-center vs Single-center	Name of Country or Registry	Inclusion Criteria (Copy and Paste)	Recruitment Time Frame (Years)
Allen	2010	MC	UNOS	Adult (aged >17 years) primary LTx patients	1987-2008
Allen	2010	MC	UNOS	Adult recipients, first transplant	1998-2008
Allyn	2016	SC	USA	All lung transplant recipients	2000-2013
Arnaoutakis	2011	SC	ISA	Patients transplanted post-LAS era only	2005-2010
Awori	2016	MC	UNOS	Adult recipients (age >17 years)	2000-2012
Bonser	2012	MC	UK	First adult lung-only from donors after brain death	1999-2010
Borro	2016	SC	Spain	All lung transplant recipients	2000-2012
Cassivi	2002	SC	ISA	Adult patients undergoing transplant for emphysema	1988-2000
Castleberry	2013	SC	USA	Adult recipients (age >17 years)	1997-2010
Ceriana	2002	SC	Italy	Single lung transplants affected by an advanced parenchymal or vascular lung disease. Age >65 years.	1992-1999
Ceron Navarro	2013	SC	Spain	Patients with COPD	1991-2008
Chaikriangkrai	2015	SC	USA	Patients receiving first transplant	2007-2013
Chaikriangkrai	2015	SC	USA	Patients receiving first transplant	2007-2013
Chandrashekaran	2015	MC	USA	Adult (aged >17 years) primary LTx patients	2000-2010
Christie	2005	MC	UNOS/ISHLT	All lung transplant recipients	1994-2000
D'Angelo	2016	SC	USA	All lung transplant recipients	2008-2013
Demir	2015	SC	Belgium	Patients who underwent either single or bilateral LTx at their centre during study period	1991-2009
DerHovanessian	2016	MC	USA	Adult (aged >17 years) primary LTx patients	1998-2012
George	2012	MC	UNOS	Adult (aged >17 years) primary LTx patients	2001-2010
Hayanga	2015	MC	UNOS	Adult (aged >17 years) primary LTx patients	2005-2012
Hayes	2016	MC	UNOS	Adult (aged >17 years) primary LTx patients	2010-2015
Hayes	2015	MC	UNOS	Adult (aged >17 years) primary LTx patients	1987-2013
Hayes	2015	MC	UNOS	All lung transplant recipients	1987-2013
Henri	2012	SC	Canada	All lung transplant recipients	1996-2006
Huppmann	2012	SC	Germany	All lung transplant recipients	1996-2007
Inci	2015	SC	Switzerland	Adult patients undergoing transplant for emphysema	1992-2013
Jacques	2012	SC	Canada	All lung transplant recipients	1997-2004
Julliard	2016	SC	USA	All patients undergoing single LTx	1999-2013
Loor	2017	SC	USA	All lung transplant recipients	1986-2016
Mason	2009	SC	USA	Adult (aged >17 years) primary LTx patients	1990-2006
Minambres	2009	SC	Spain	Patients diagnosed with idiopathic emphysema or alpha-1-antitrypsin-deficiency-related emphysema	1997-2011
Minambres	2010	SC	Spain	Adult (>55 years) primary LTx patients	1997-2010
Mollberg	2015	SC	USA	Adult (aged >17 years) primary LTx patients	2000-2010
Neurohr	2010	SC	Germany	Patients with idiopathic pulmonary fibrosis	1997-2008
Newton	2017	SC	USA	Patients with pulmonary fibrosis	2007-2014
Plantier	2010	SC	France	All lung transplant recipients	1988-2007
Schaffer	2015	MC	UNOS	Patients with pulmonary fibrosis or COPD	2005-2012
Shigemura	2013	SC	USA	All lung transplant recipients	2004-2010
Shigemura	2014	SC	USA	Primary LTx patients	2004-2011
Shigemura	2013	SC	USA	All lung transplant recipients	2004-2008
Shino	2013	SC	USA	All lung transplant recipients	2000-2010
Taimeh	2016	MC	UNOS	All lung transplant recipients	1987-2012
Vadnerkar	2010	SC	USA	All lung transplant recipients	2006-2008
Wehbe	2012	SC	USA	All lung transplant recipients	1997-2009
Whitson	2014	MC	UNOS	Adult (aged >17 years) primary LTx patients with deceased donors	1987-2010
Xue	2014	SC	China	All lung transplant recipients	2002-2011
Zalunardo	2010	SC	Switzerland	All lung transplant recipients	1996-2006

SC – single center; MC – Multi-center; UNOS – United Network for Organ Sharing; LTx – Lung Transplant; LAS – Lung allocation score; COPD – Chronic Obstructive Pulmonary Disease;

APPENDIX A Continued – Characteristics and demographics of included studies

First Author	Year	Follow-up	n	Female Recipient	Female Donor	Recipient Age	Donor Age	Recipient Black Race
Allen	2010	3.1 ± 2.7	11411	46%	NR	32 ± 14	33 ± 14	7%
Allen	2010	3.1 ± 2.7	11385	46%	NR	NR	NR	7%
Allyn	2016	2.7 ± 2	563	42%	NR	NR	NR	NR
Arnaoutakis	2011	1.0 ± 0.8	106	49%	NR	NR	NR	17%
Awori	2016	NR	16156	44%	39%	31 (21 - 45)*	31 (21 - 45)*	NR
Bonser	2012	NR	1295	42%	52%	42 (28 - 51)*	42 (28 - 51)*	NR
Borro	2016	NR	73	15%	NR	NR	NR	NR
Cassivi	2002	3.7 ± 3.3	306	53%	NR	NR	NR	NR
Castleberry	2013	4.4 ± NR	791	42%	NR	32 (42 - 68)*	32 (21 - 47)*	NR
Ceriana	2002	NR	66	27%	NR	49 (18 - 64)***	NR	NR
Ceron Navarro	2013	NR	107	12%	NR	53 ± 8	35 ± 14	NR
Chaikriangkrai	2015	2.3 ± 1.4	293	43%	NR	57 ± 13	NR	NR
Chaikriangkrai	2015	2.3 (1.0 - 3.6)*	324	42%	NR	57 ± 13	NR	6%
Chandrashekaran	2015	NR	355	45%	NR	59 (52 - 65)*	NR	NR
Christie	2005	NR	5262	NR	NR	NR	NR	NR
D'Angelo	2016	NR	652	42%	NR	61 (50 - 67)*	NR	6%
Demir	2015	5.4 ± 3.8	461	44%	47%	49 ± 13	40 ± 14	NR
DerHovanessian	2016	4.2 (2.6 - 6.8)*	389	36%	NR	57 (46 - 63)*	NR	9%
George	2012	NR	11878	46%	40%	52 ± 13	33 ± 14	8%
Hayanga	2015	NR	10304	41%	NR	55 ± 13	34 ± 4	NR
Hayes	2016	NR	8228	39%	40%	55 ± 13	NR	8%
Hayes	2015	NR	23582	45%	39%	52 ± 13	NR	7%
Hayes	2015	NR	23905	45%	38%	NR	NR	7%
Henri	2012	NR	224	50%	NR	47 ± 13	NR	NR
Huppmann	2012	2.9 ± 0.2	206	50%	49%	48 ± 1	NR	NR
Inci	2015	3.0 (1.2 - 8.3)*	108	46%	34%	56 (51 - 60)*	50 (37 - 57)*	NR
Jacques	2012	3 (2.2)**	174	52%	41%	46 ± 14	35 ± 14	NR
Julliard	2016	NR	279	28%	NR	NR	NR	NR
Loor	2017	NR	876	52%	40%	50 ± 13	NR	NR
Mason	2009	4.5 ± 3.1	469	49%	50%	48 ± 12	36 ± 15	NR
Minambres	2009	NR	92	16%	26%	55 ± 7	37 ± 12	NR
Minambres	2010	NR	33	48%	52%	54 ± 9	57 ± 2	NR
Mollberg	2015	NR	452	46%	NR	52 ± 13	NR	NR
Neurohr	2010	3.2 ± 0.2	76	43%	NR	52 ± 1	NR	NR
Newton	2017	5.0 ± 2.5	82	30%	NR	59 ± 9	NR	6%
Plantier	2010	1.1 (3.6)**	258	31%	NR	51 ± 10	NR	NR
Schaffer	2015	2.0 (0.7 - 3.9)*	11892	61%	NR	60 ± 7	35 ± 14	4%
Shigemura	2013	0 - 5	759	52%	43%	51 ± 9	43 ± 19	NR
Shigemura	2014	NR	873	45%	NR	54 ± 9	NR	NR
Shigemura	2013	0 - 5	293	52%	43%	55 ± 21	55 ± 9	NR
Shino	2013	NR	441	41%	NR	61 ± NR	NR	NR
Taimeh	2016	2.8 (0 - 22.3)*	20971	47%	39%	50 ± 15	32 ± 14	NR
Vadnerkar	2010	NR	121	41%	47%	67 (63 - 75)***	44 (14 - 71)***	NR
Wehbe	2012	2.2 (0.4 - 6.9)*	657	42%	NR	53 ± 12	NR	NR
Whitson	2014	NR	18250	47%	38%	51 ± 13	32 ± 14	NR
Xue	2014	2.1 ± 2.0	88	24%	NR	53 ± 14	NR	NR
Zalunardo	2010	NR	169	46%	NR	42 ± 16	NR	NR

NR – Not reported

* median (25^{th} – 75^{th} %)
** median (IQR)
*** median (range)

APPENDIX A Continued – Characteristics and demographics of included studies

First Author	Year	Recipient BMI	Single Lung Transplant	Ischemic Time Overall	Diabetes mellitus	Hypertension	Obstructive	Restrictive
Allen	2010	24 ± 2	46%	4.8 ± 2.0	12%	15%	37%	25%
Allen	2010	24 ± 5	46%	4.8 ± 2.0	12%	15%	37%	25%
Allyn	2016	NR	47%	5.0 ± 1.3	NR	3%	26%	56%
Arnaoutakis	2011	24 ± 5	12%	NR	NR	NR	NR	NR
Awori	2016	NR	39%	5.0 ± 1.7	15%	3%	32%	NR
Bonser	2012	22 (19 - 26)*	30%	4.7 (3.8 - 5.5)*	13%	NR	37%	20%
Borro	2016	NR	55%	NR	NR	NR	75%	NR
Cassivi	2002	NR	72%	5.9 ± 1.3	NR	NR	72%	NR
Castleberry	2013	24 (20 - 28)*	17%	6.2 (4.5 - 9.6)*	12%	24%	42%	39%
Ceriana	2002	24 ± 7	100%	3.9 (3.1 - 5.1)*	NR	NR	NR	NR
Ceron Navarro	2013	23 ± 2	29%	NR	NR	NR	NR	NR
Chaikriangkrai	2015	26 ± 6	37%	3.4 ± 1.1	29%	54%	26%	65%
Chaikriangkrai	2015	26 ± 5	38%	3.4 ± 1.1	NR	NR	27%	64%
Chandrashekaran	2015	25 (21 - 28)*	52%	3.8 ± NE	NR	NR	46%	49%
Christie	2005	NR	NR	NR	NR	NR	NR	NR
D'Angelo	2016	NR	15%	5.6 (4.8 - 6.4)*	21%	37%	35%	47%
Demir	2015	NR	32%	NR	NR	NR	42%	18%
DerHovanessian	2016	NR	24%	NR	NR	NR	80%	84%
George	2012	25 ± 7	42%	4.9 ± 1.7	14%	13%	35%	28%
Hayanga	2015	25 ± 5	34%	5.1 ± 1.7	18%	NR	28%	36%
Hayes	2016	25 ± 4	38%	5.2 ± 1.7	19%	NR	24%	45%
Hayes	2015	24 ± 5	NR	4.8 ± 1.7	NR	3%	34%	30%
Hayes	2015	24 ± 5	45%	4.8 ± 1.7	NR	NR	NR	NR
Henri	2012	NR	44%	NR	NR	17%	NR	19%
Huppmann	2012	NR	44%	5.8 ± 6.0	NR	NR	NR	31%
Inci	2015	21 (19 - 24)*	10%	NR	NR	NR	72%	NR
Jacques	2012	23 ± 5	51%	NR	13%	13%	43%	14%
Julliard	2016	26 ± NR	100%	NR	26%	38%	35%	41%
Loor	2017	NR	56%	NR	NR	NR	46%	22%
Mason	2009	24 ± 5	58%	3.7 ± 1.2	NR	NR	67%	17%
Minambres	2009	24 ± 4	22%	5.5 ± 3.5	NR	NR	NR	NR
Minambres	2010	24 ± 4	18%	5.4 ± 1.4	NR	NR	NR	NR
Mollberg	2015	25 ± 4	24%	NR	15%	NR	32%	NR
Neurohr	2010	25 ± 1	61%	5.8 ± 0.3	NR	NR	NR	NR
Newton	2017	NR	15%	NR	NR	NR	NR	100%
Plantier	2010	22 ± 5	73%	NR	5%	12%	59%	26%
Schaffer	2015	26 ± 4	35%	4.9 ± 1.5	10%	NR	27%	35%
Shigemura	2013	NR	30%	6.5 ± 2.1	NR	25%	NR	NR
Shigemura	2014	26 ± 8	82%	5.8 ± 2.1	NR	55%	2%	NR
Shigemura	2013	NR	NR	NR	NR	NR	63%	41%
Shino	2013	NR	43%	NR	NR	41%	28%	44%
Taimeh	2016	NR	43%	4.8 ± 1.7	NR	NR	NR	NR
Vadnerkar	2010	26 (17 - 37)***	47%	4.6 (2.2 - 9.6)***	18%	NR	37%	46%
Wehbe	2012	25 ± 6	43%	NR	15%	28%	35%	32%
Whitson	2014	NR	49%	4.7 (1.7)	NR	NR	37%	23%
Xue	2014	NR	57%	5.9 ± 5.8	16%	22%	30%	52%
Zalunardo	2010	NR	NR	NR	NR	NR	51%	NR

NR – Not reported

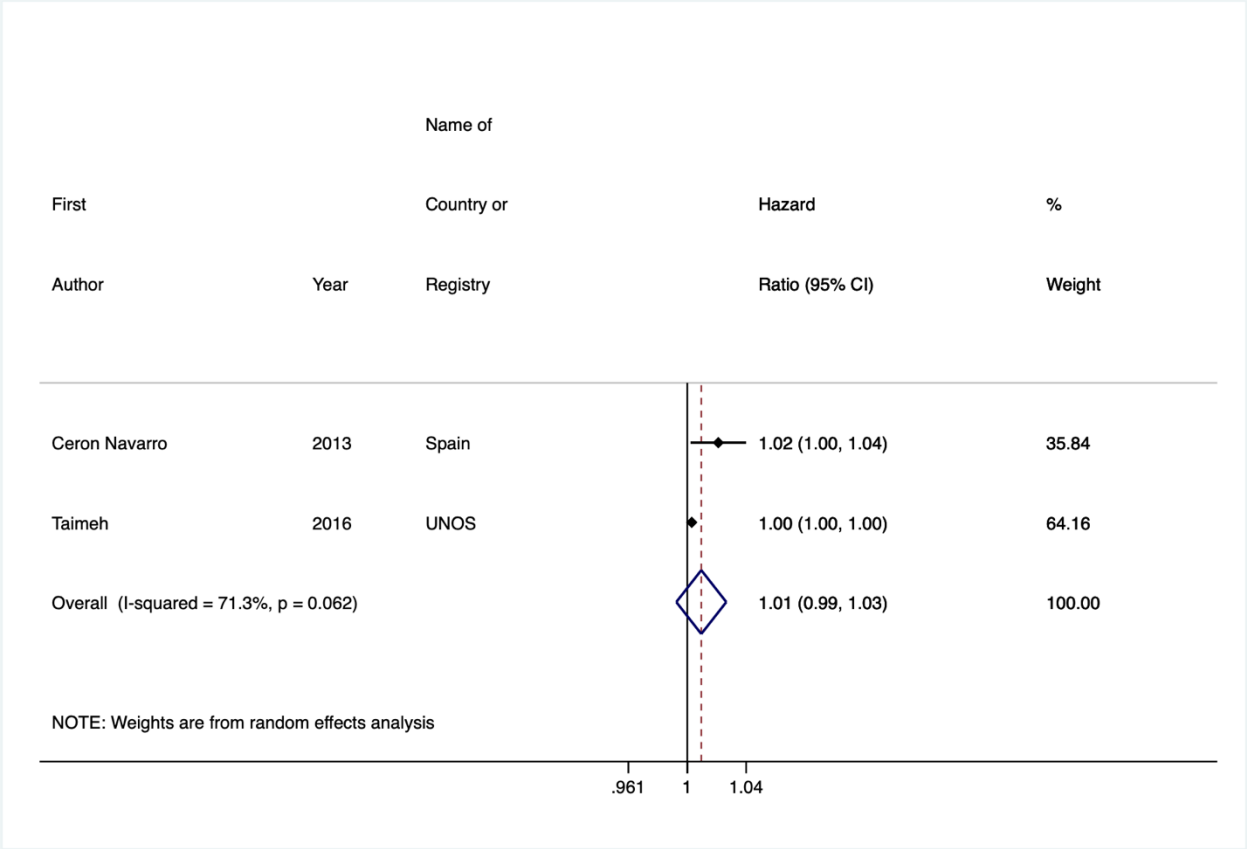
* median (25^{th} – 75^{th} %)
** median (IQR)
*** median (range)

APPENDIX B – Risk of bias of included studies

First Author	Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall
Cassivi	2002	Low	NR	Low	Low	Low	Low	Low
Ceriana	2002	Low	NR	Low	Low	Low	Low	Low
Christie	2005	Moderate	NR	Low	Low	Low	High	High
Minambres	2009	Low	NR	Low	Low	Moderate	Low	Low
Mason	2009	Low	NR	Low	Moderate	Low	High	High
Plantier	2010	Low	NR	Low	Low	High	High	High
Zalunardo	2010	Low	NR	Low	Low	High	Low	Low
Minambres	2010	Low	NR	Low	Low	Low	Low	Low
Vadnerkar	2010	Low	NR	Low	Low	Low	Low	Low
Allen	2010	Low	NR	Low	Low	Moderate	Low	Low
Allen	2010	Low	NR	Moderate	Low	Moderate	High	High
Neurohr	2010	Moderate	NR	Low	Low	Moderate	Low	High
Arnaoutakis	2011	Low	Low	Moderate	Low	Low	Low	Low
Bonser	2012	Low	Low	Low	Low	Moderate	High	High
Jacques	2012	Low	Low	Low	Low	Moderate	High	High
George	2012	Low	NR	Low	Low	Low	Low	Low
Wehbe	2012	Low	NR	Low	Low	Moderate	Moderate	High
Henri	2012	Moderate	Low	Low	Moderate	Low	Low	High
Huppmann	2012	Moderate	NR	Low	Low	Low	Low	Low
Castleberry	2013	Low	NR	Low	Low	High	Low	Low
Ceron Navarro	2013	Low	NR	Low	Low	Low	Low	Low
Shigemura	2013	Low	NR	Low	Low	Low	Low	Low
Shigemura	2013	Low	NR	Low	Low	Low	Low	Low
Shino	2013	Low	NR	Low	Low	Low	Moderate	Low
Shigemura	2014	Low	Low	Low	Low	Low	Low	Low
Xue	2014	Low	NR	Moderate	Low	Moderate	High	High
Whitson	2014	Moderate	Low	Low	Low	Low	High	High
Schaffer	2015	Low	Low	Low	Low	Low	High	Low
Chandrashekaran	2015	Low	Low	Low	Low	Moderate	High	High
Inci	2015	Low	Low	Moderate	Low	Low	Low	Low
Chaikriangkrai	2015	Low	NR	Low	Low	High	Low	Low
Hayes	2015	Low	NR	Low	Low	Low	High	Low
Demir	2015	Low	NR	Low	Low	Low	Low	Low
Mollberg	2015	Low	NR	Low	Low	Low	Low	Low
Hayanga	2015	Low	NR	Low	Low	Moderate	High	High
Hayes	2015	Low	NR	Low	Low	Moderate	High	High
Chaikriangkrai	2015	Low	NR	Low	Low	Moderate	Low	Low
D'Angelo	2016	Low	Low	Low	Low	Low	High	Low
Allyn	2016	Low	Low	Low	Low	Low	Low	Low
DerHovanessian	2016	Low	Low	Low	Low	Low	Low	Low
Awori	2016	Low	NR	Low	Low	Low	Low	Low
Borro	2016	Low	NR	Low	Low	Low	Low	Low
Hayes	2016	Low	NR	Low	Low	Low	Low	Low
Julliard	2016	Low	NR	Low	Low	Moderate	Low	Low
Taimeh	2016	Low	NR	Low	Low	Moderate	Low	Low
Loor	2017	Low	NR	Low	Low	High	Low	Low
Newton	2017	Low	NR	Low	Low	Low	Low	Low

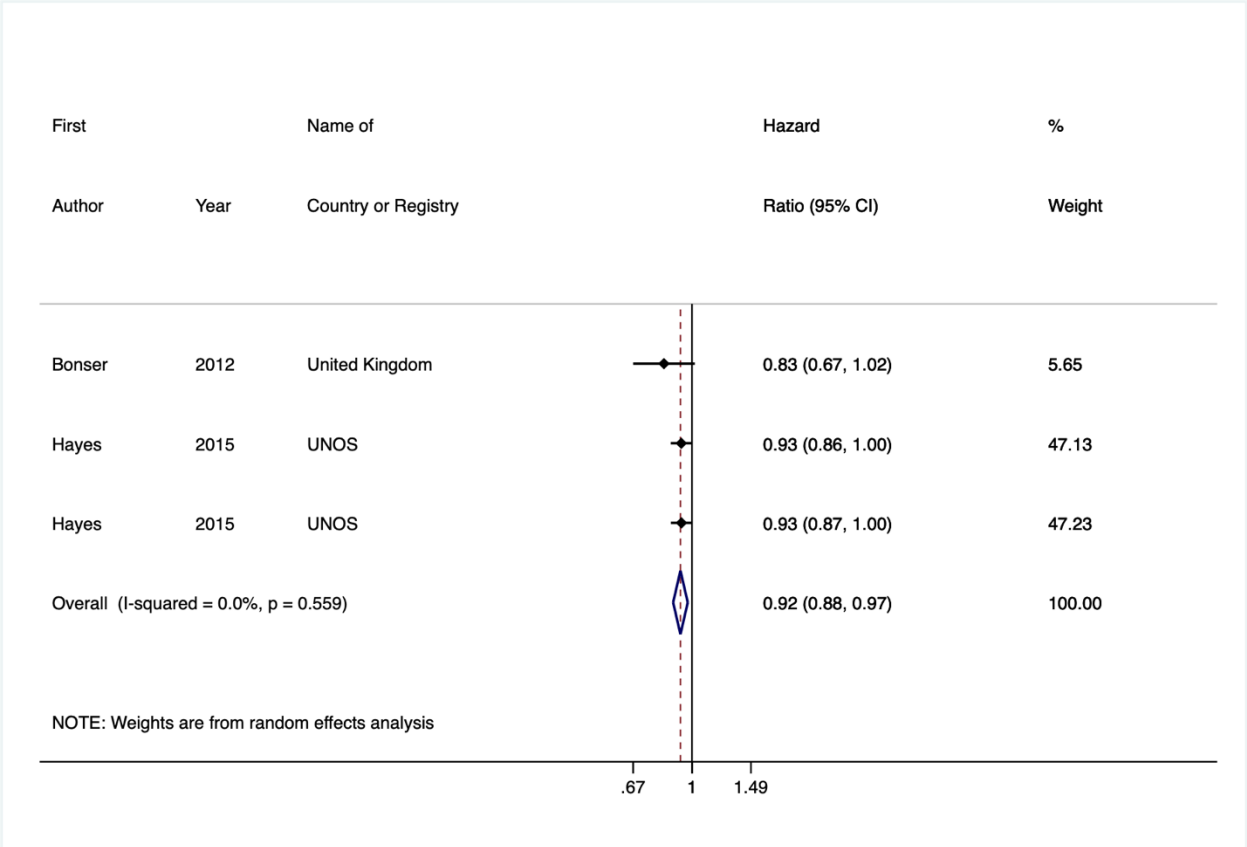
NR – Not reported

APPENDIX C – Meta-analysis of donor age



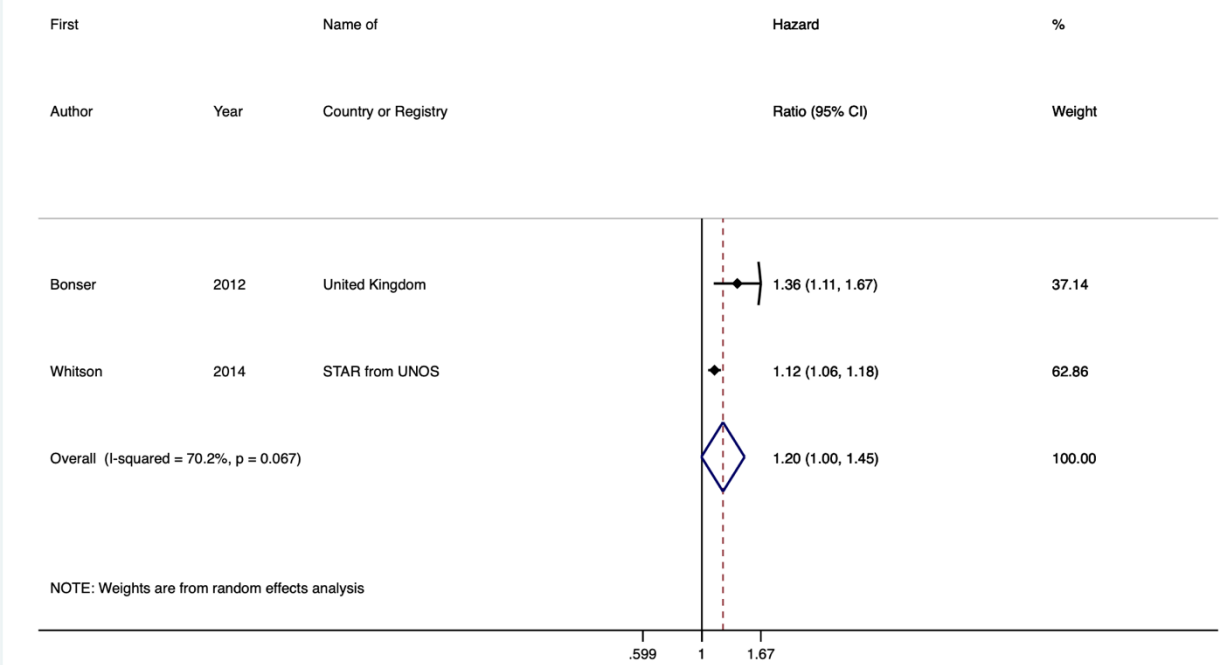
HR per 10-year increase

APPENDIX D – Meta-analysis of donor sex

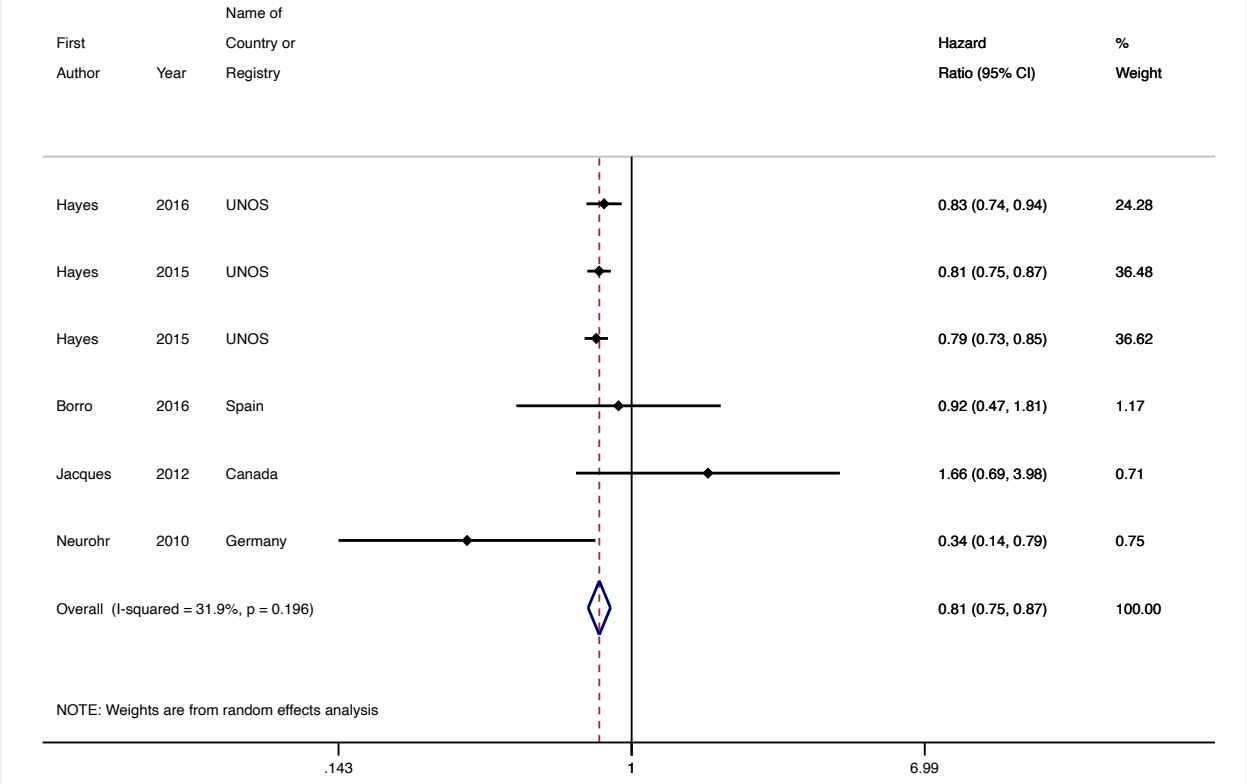


Male sex compared to female sex

APPENDIX E – Meta-analysis of donor smoking

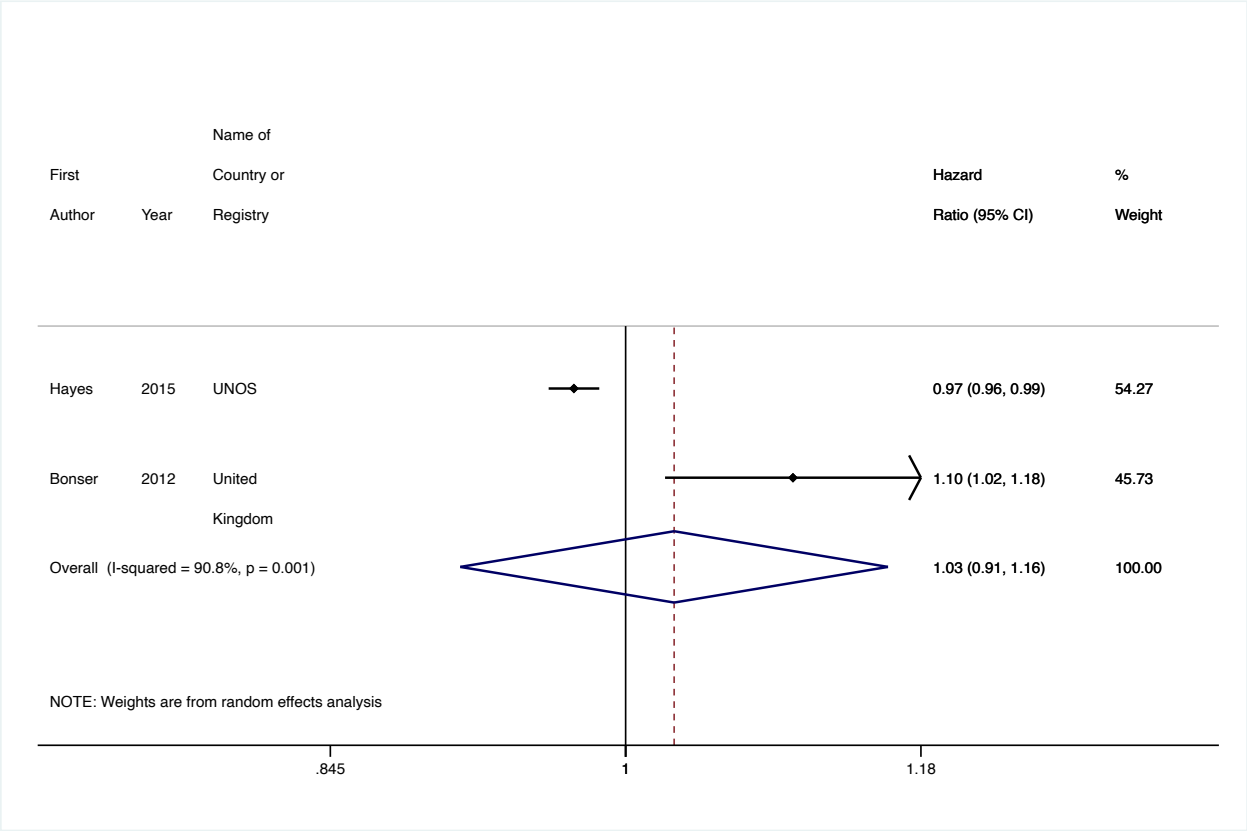


APPENDIX F – Meta-analysis of type of transplant



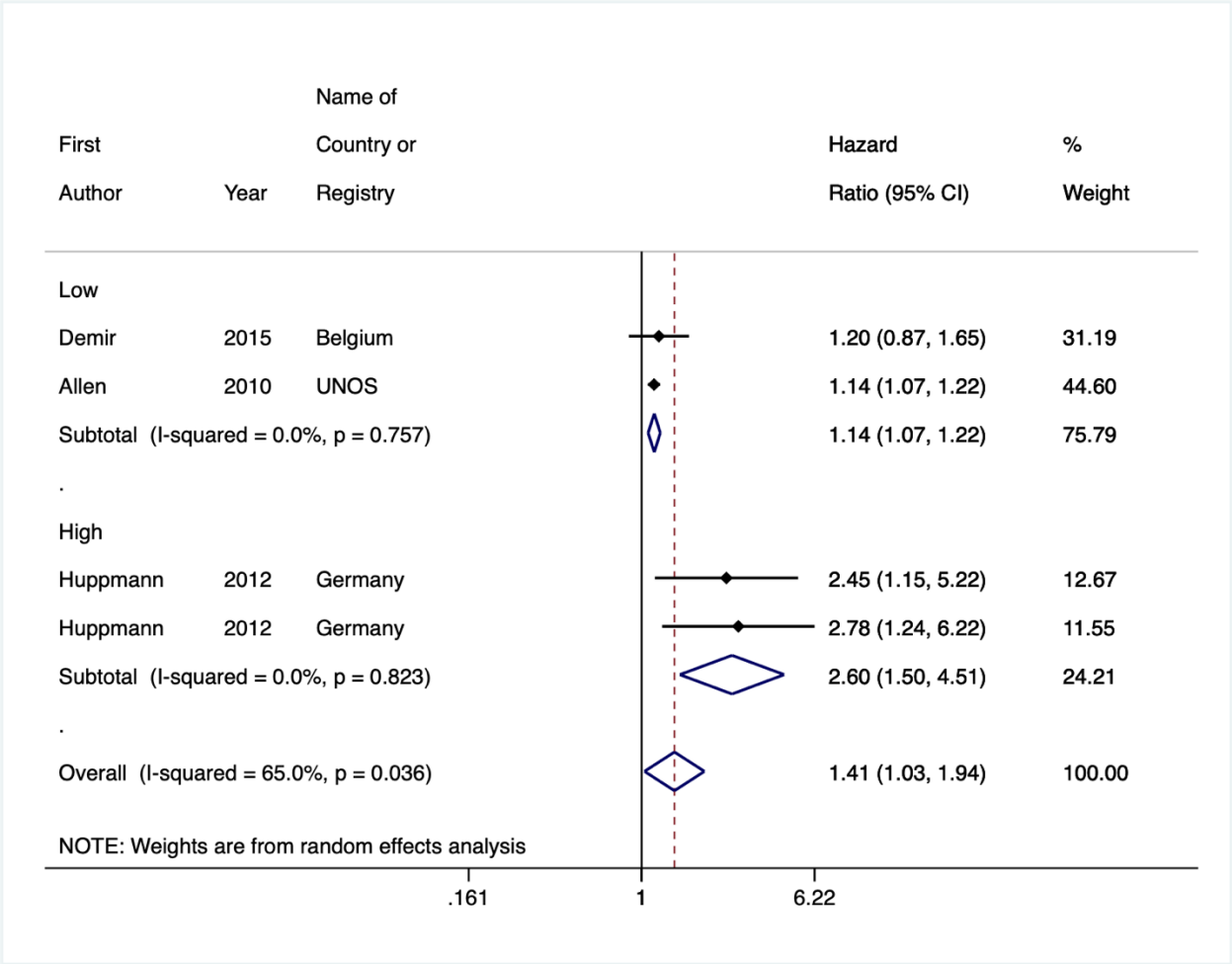
Bilateral transplant vs single lung transplant

APPENDIX G – Meta-analysis of ischemic time



HR per 1-hour increase

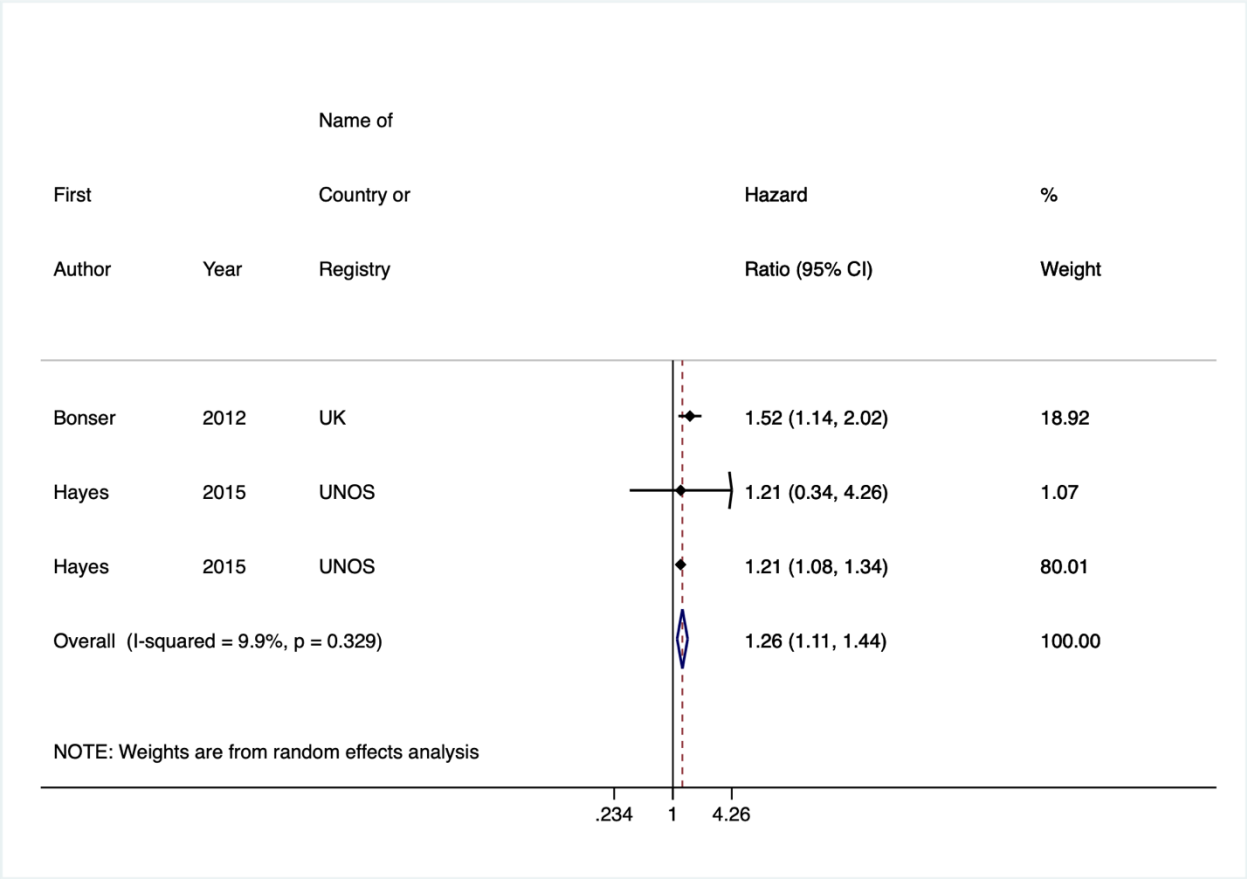
APPENDIX H – Meta-analysis of CMV mismatch



Any mismatch compared to no-mismatch

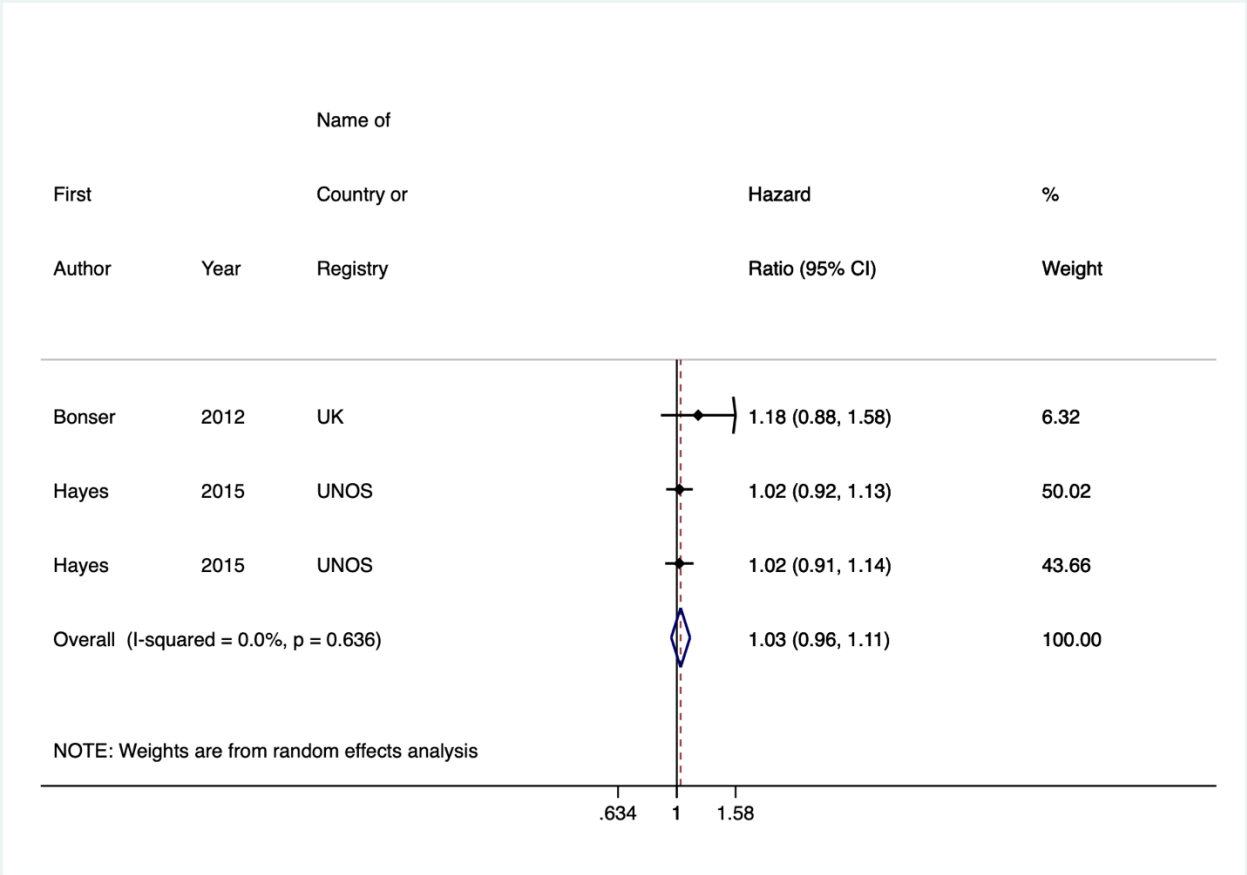
Low = low risk of bias. High = high risk of bias. Due to significant difference between low and high risk of bias studies, only low risk of bias studies inform the final estimate in table 1.

APPENDIX I – Meta-analysis of CMV Status (R- vs D+)



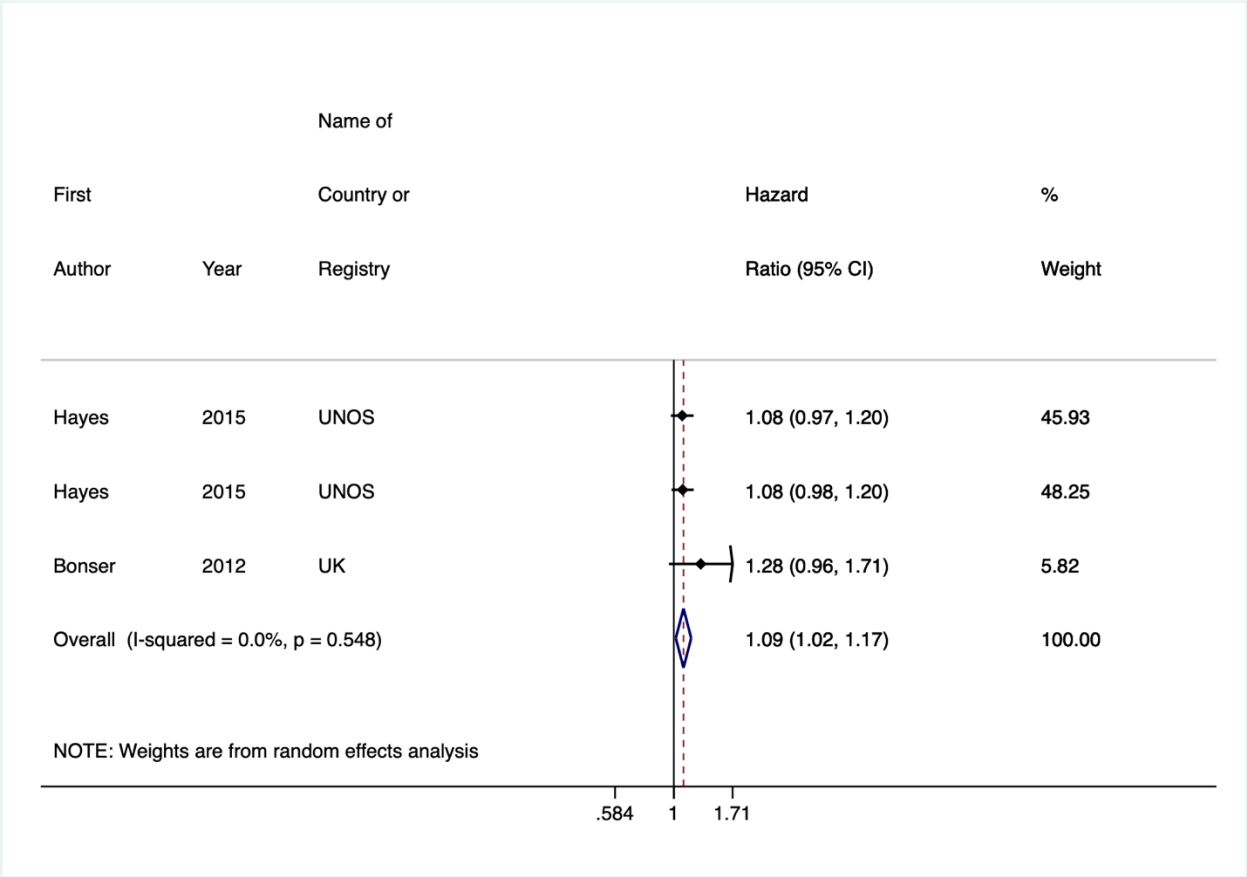
Recipient negative and donor positive compared to recipient negative and donor negative

APPENDIX J – Meta-analysis of CMV Status (R+ vs D-)



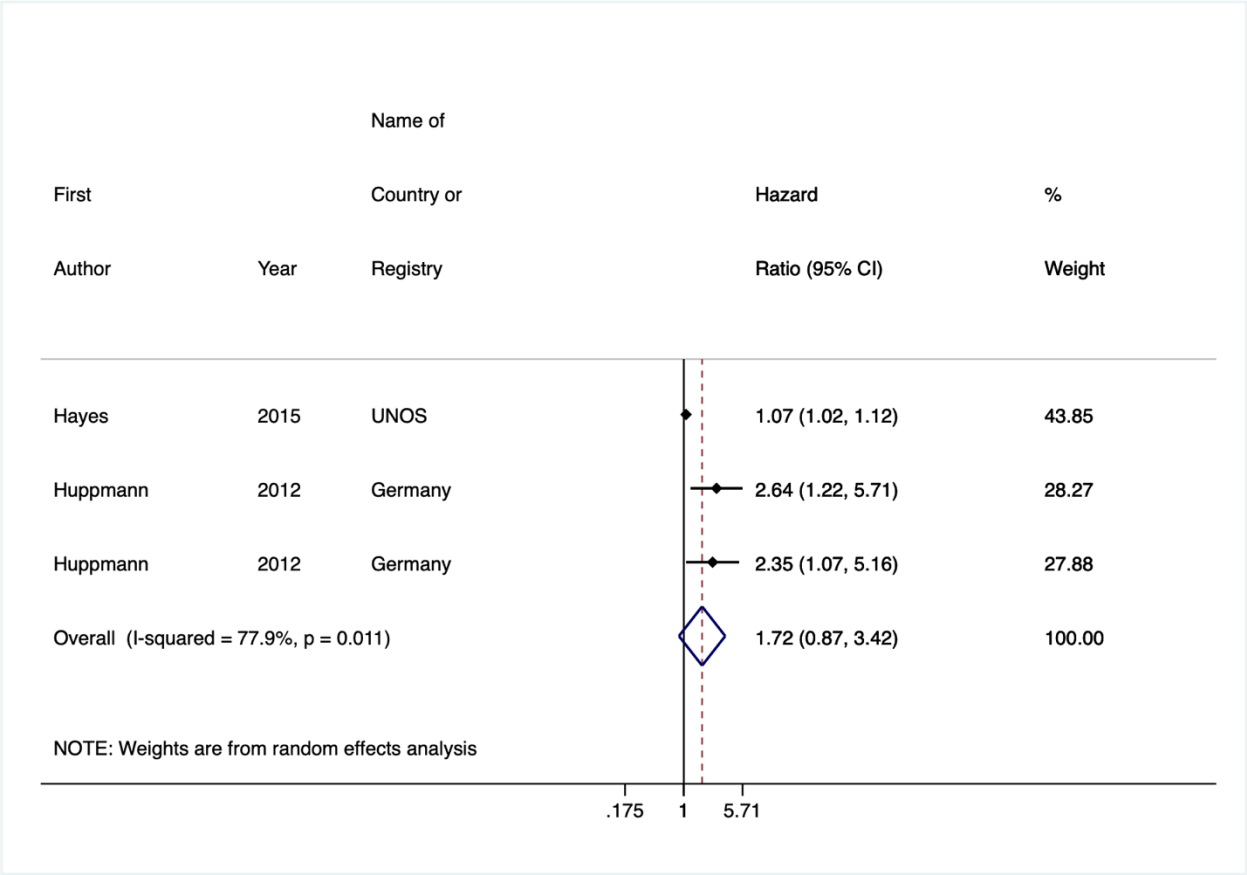
Recipient positive and donor negative compared to recipient negative and donor negative

APPENDIX K – Meta-analysis of CMV Status (R+ vs D+)



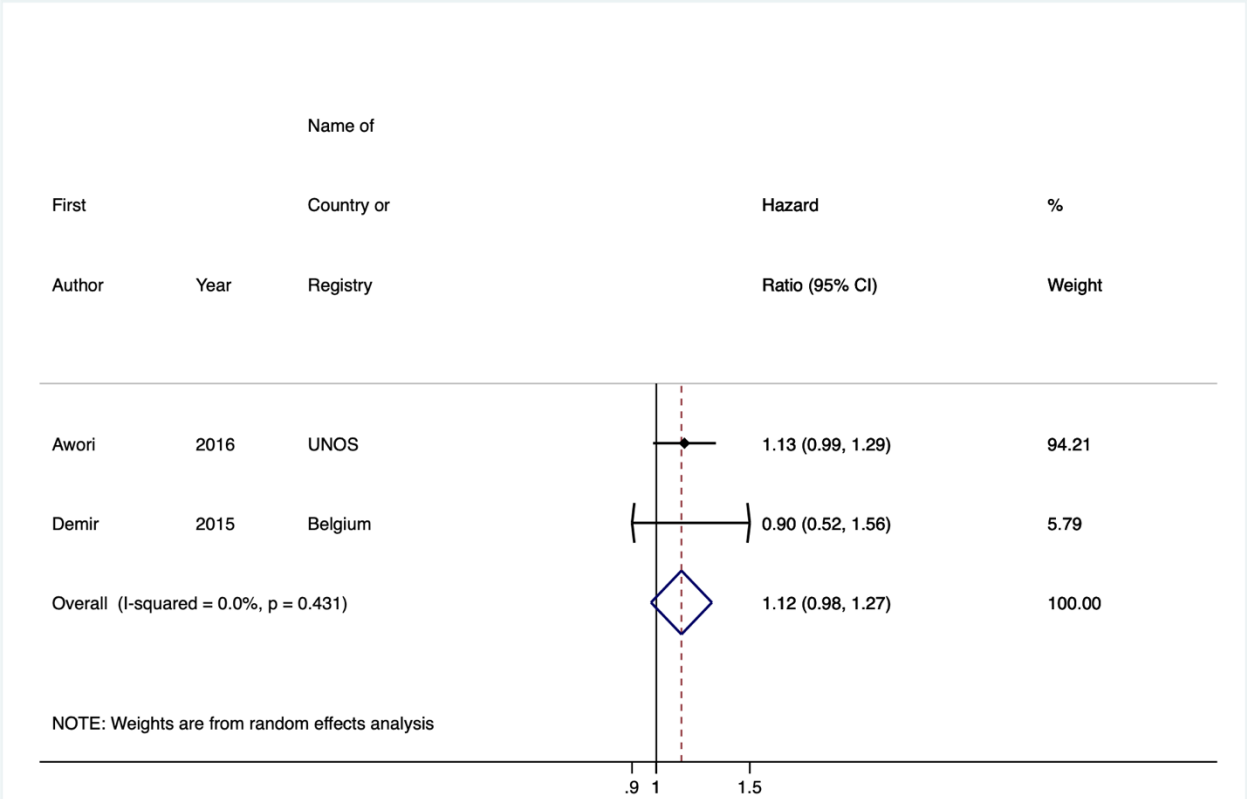
Recipient positive and donor positive compared to recipient negative and donor negative

APPENDIX L – Meta-analysis of HLA-A locus 2 mismatch

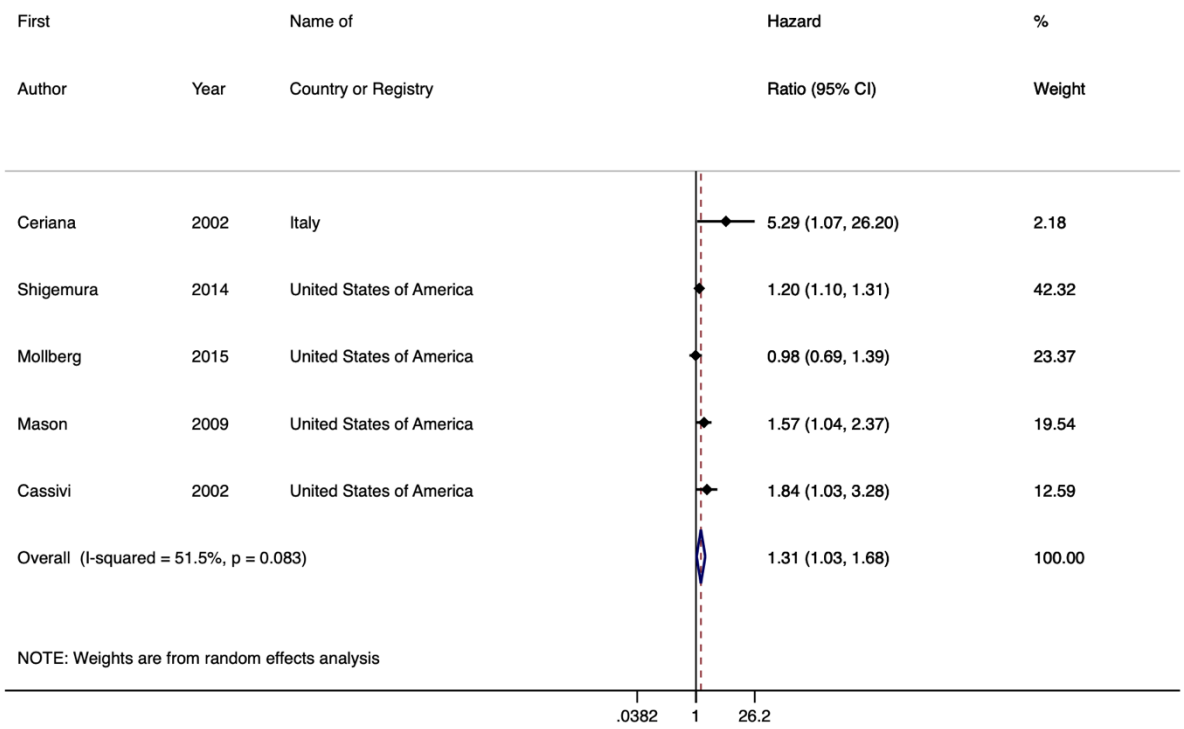


Mismatch compared to no-mismatch

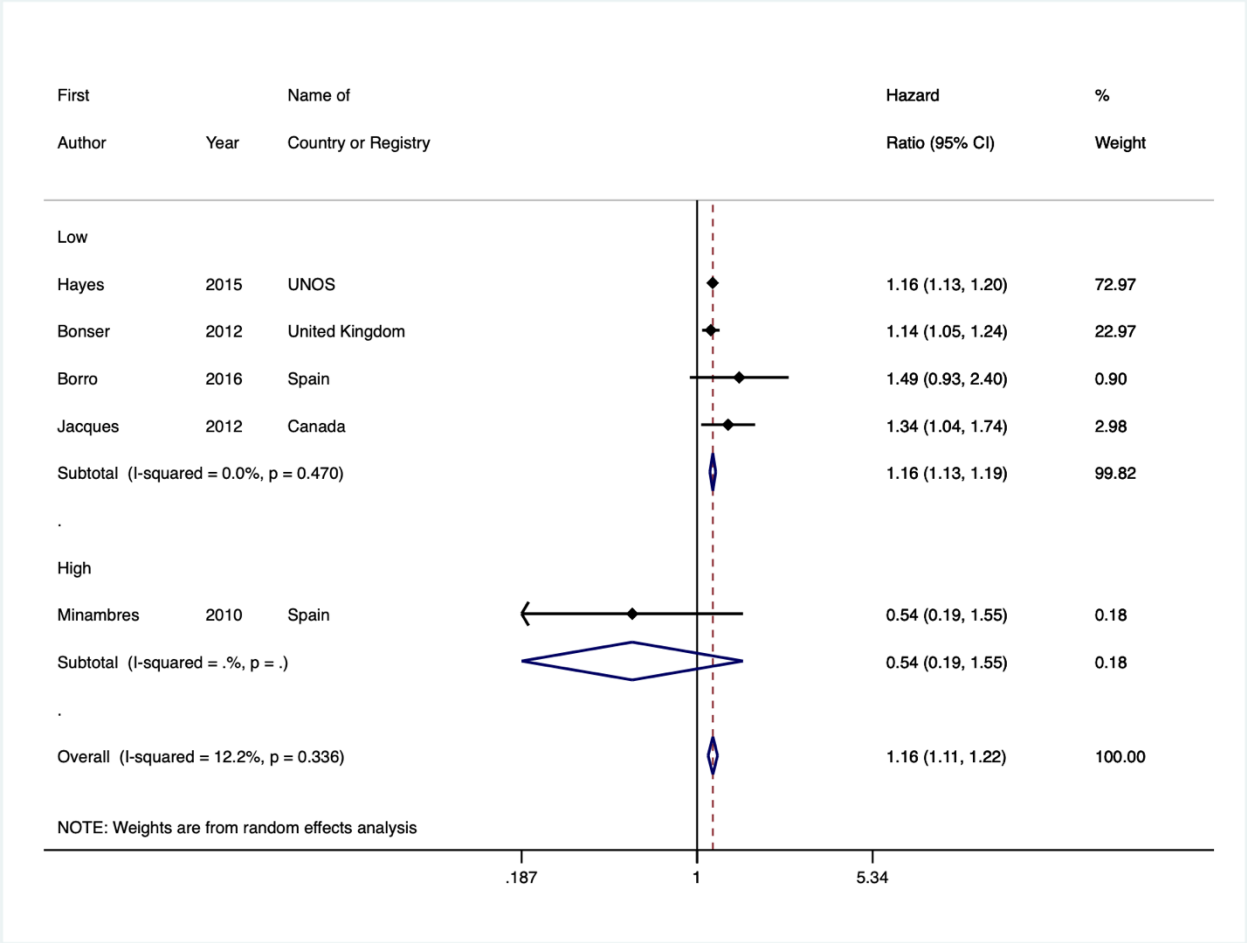
APPENDIX M – Meta-analysis of ABO incompatibility



APPENDIX N – Meta-analysis of cardio-pulmonary bypass use

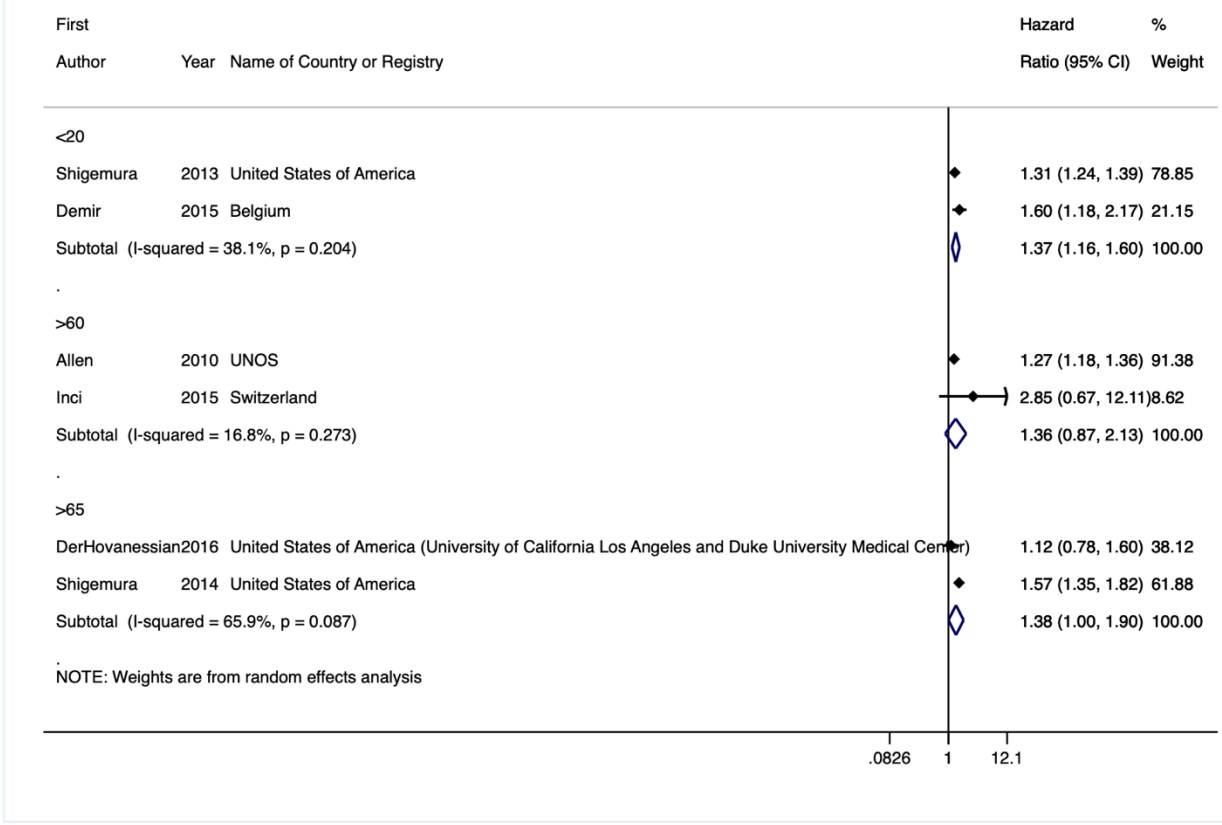


APPENDIX O – Meta-analysis of recipient age (continuous)

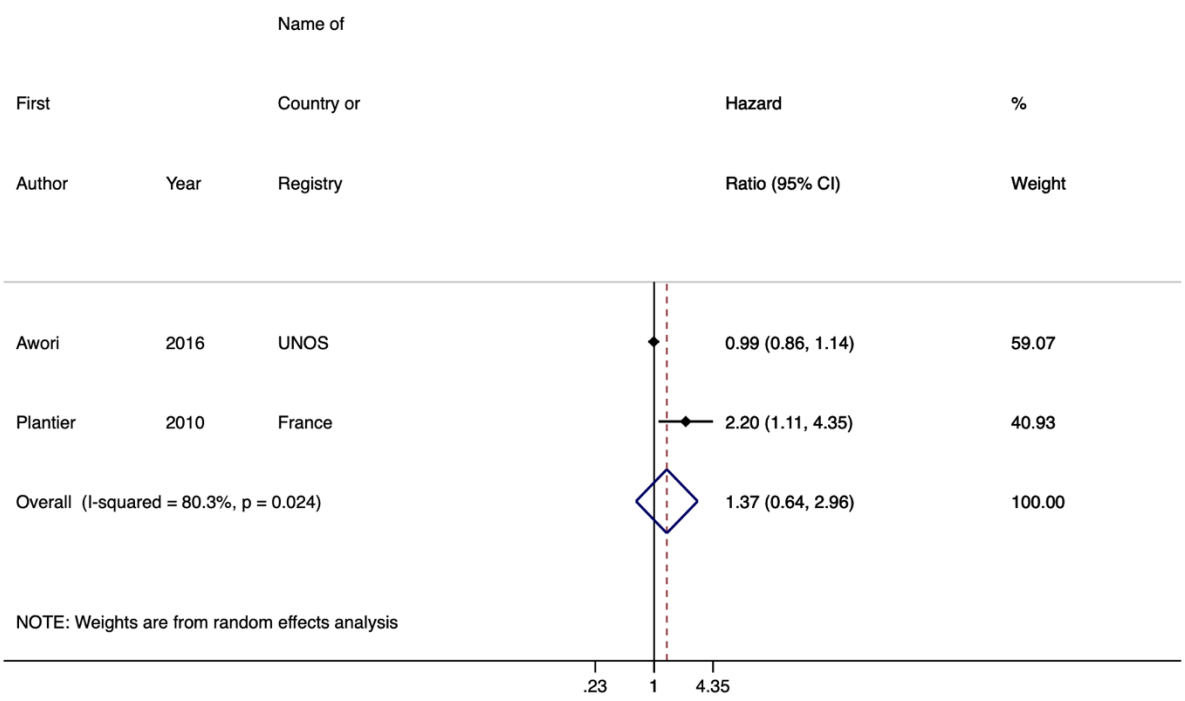


HR per 10-year increase. Low = low risk of bias. High = high risk of bias. Due to significant difference between low and high risk of bias studies, only low risk of bias studies inform the final estimate in table 1.

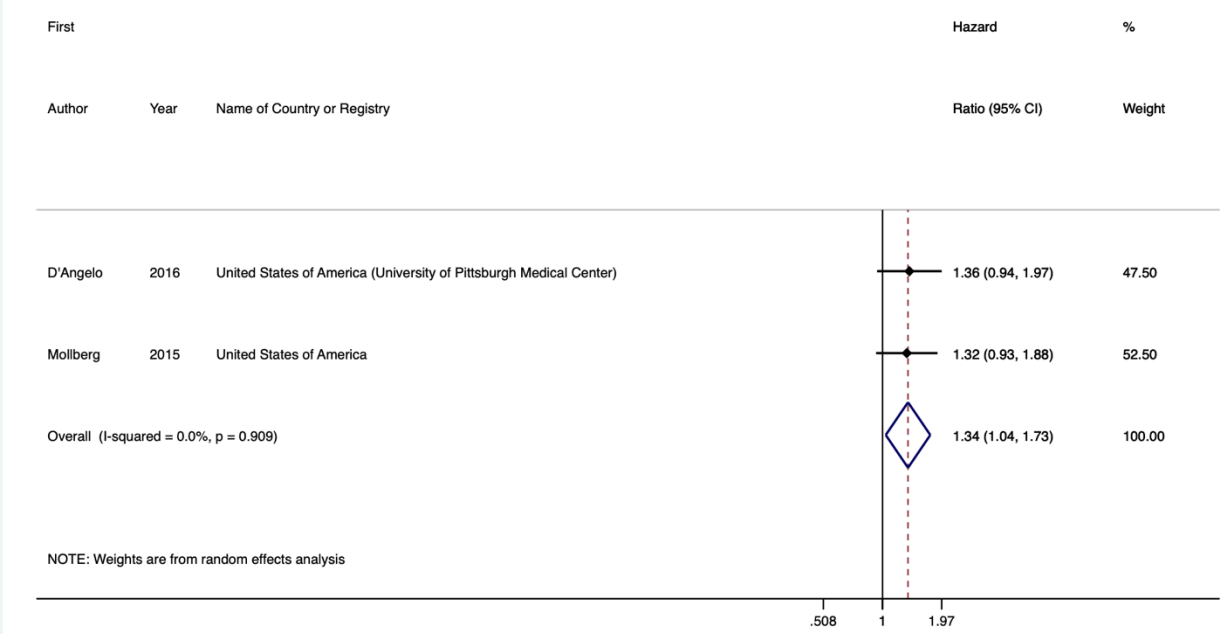
APPENDIX P – Meta-analysis of recipient age (categorical)



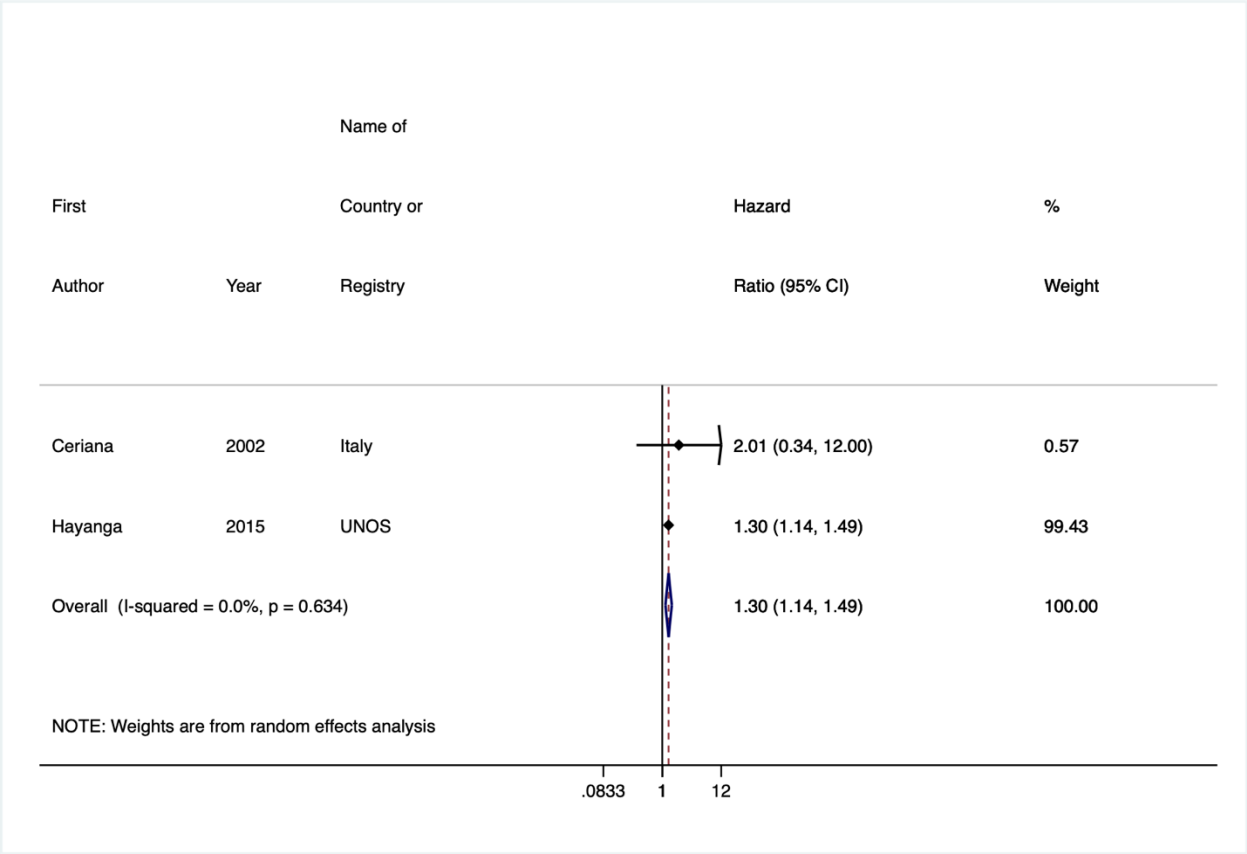
APPENDIX Q – Meta-analysis of recipient diabetes



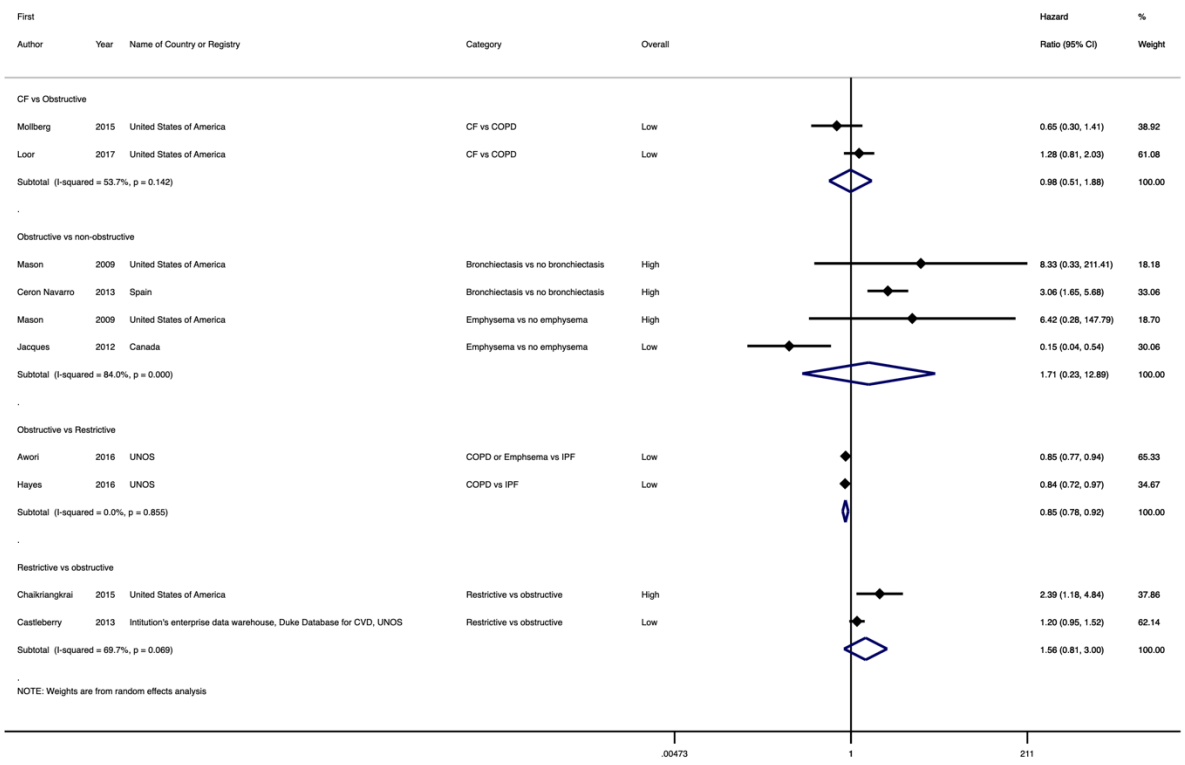
APPENDIX R – Meta-analysis of recipient hypertension



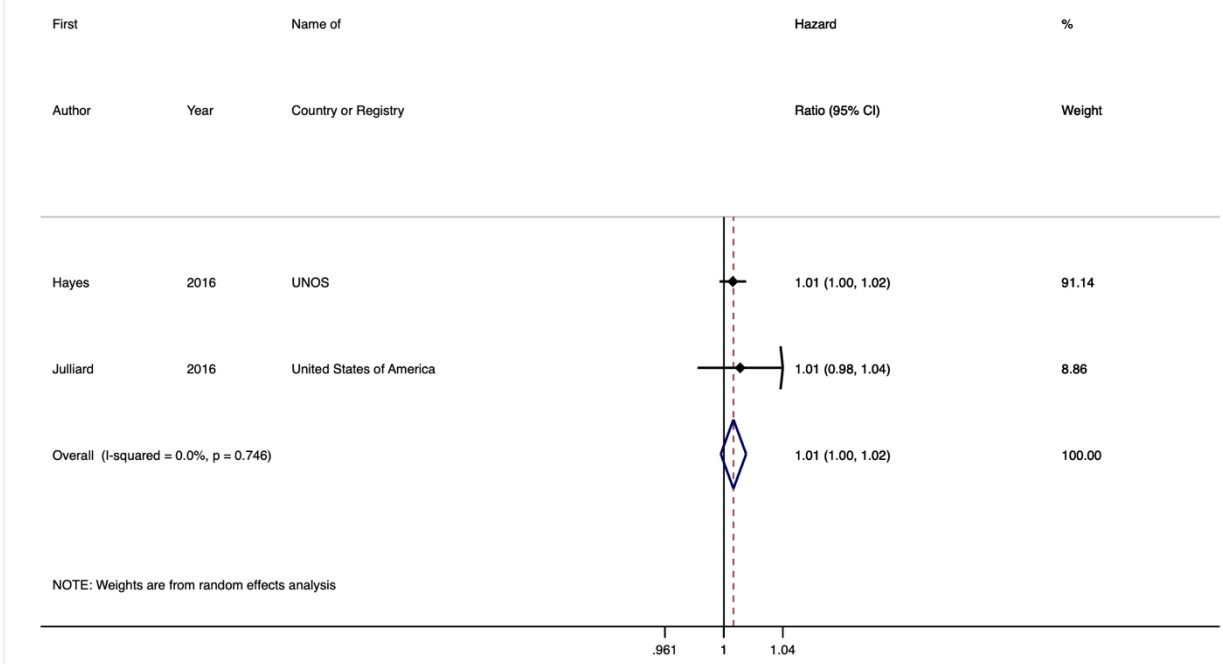
APPENDIX S – Meta-analysis of recipient BMI



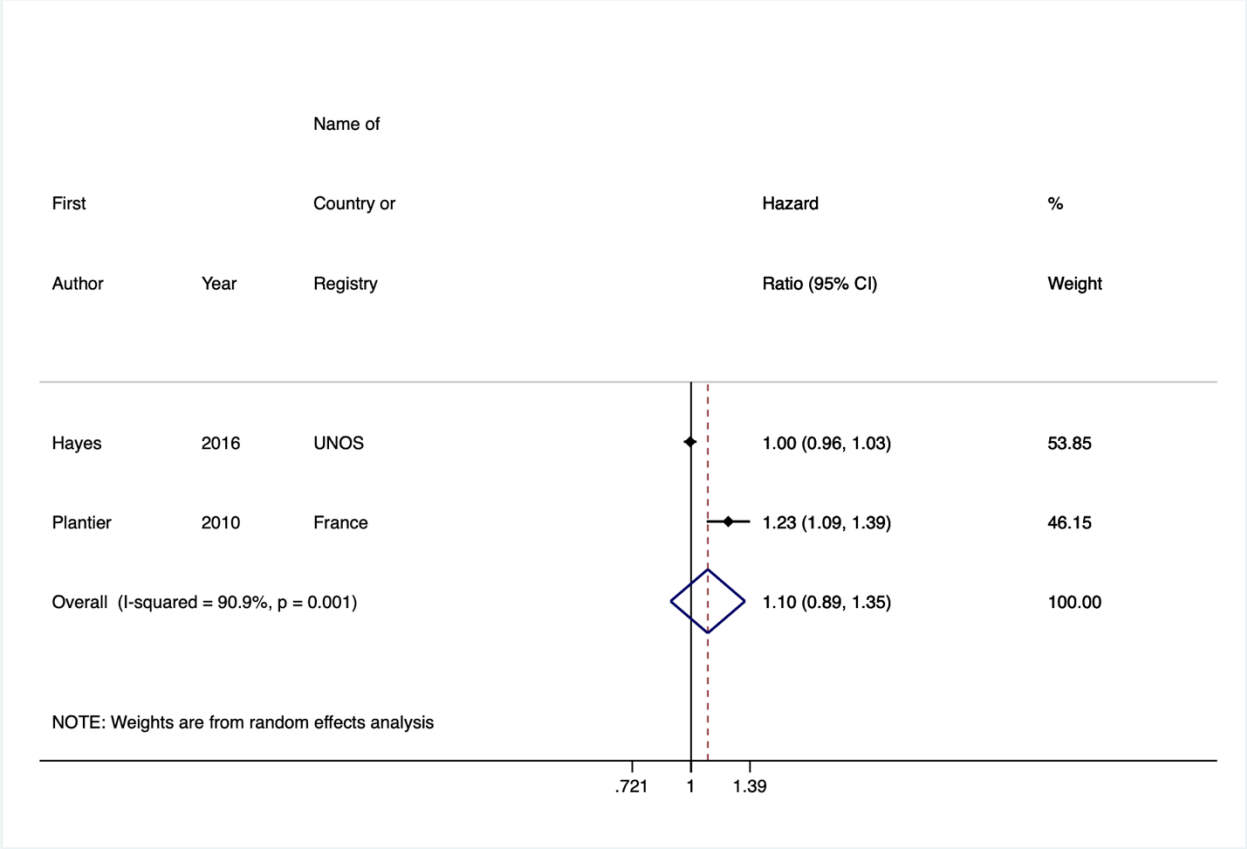
APPENDIX T – Meta-analysis of recipient aetiology



APPENDIX U – Meta-analysis of recipient pre-transplant wedge pressure

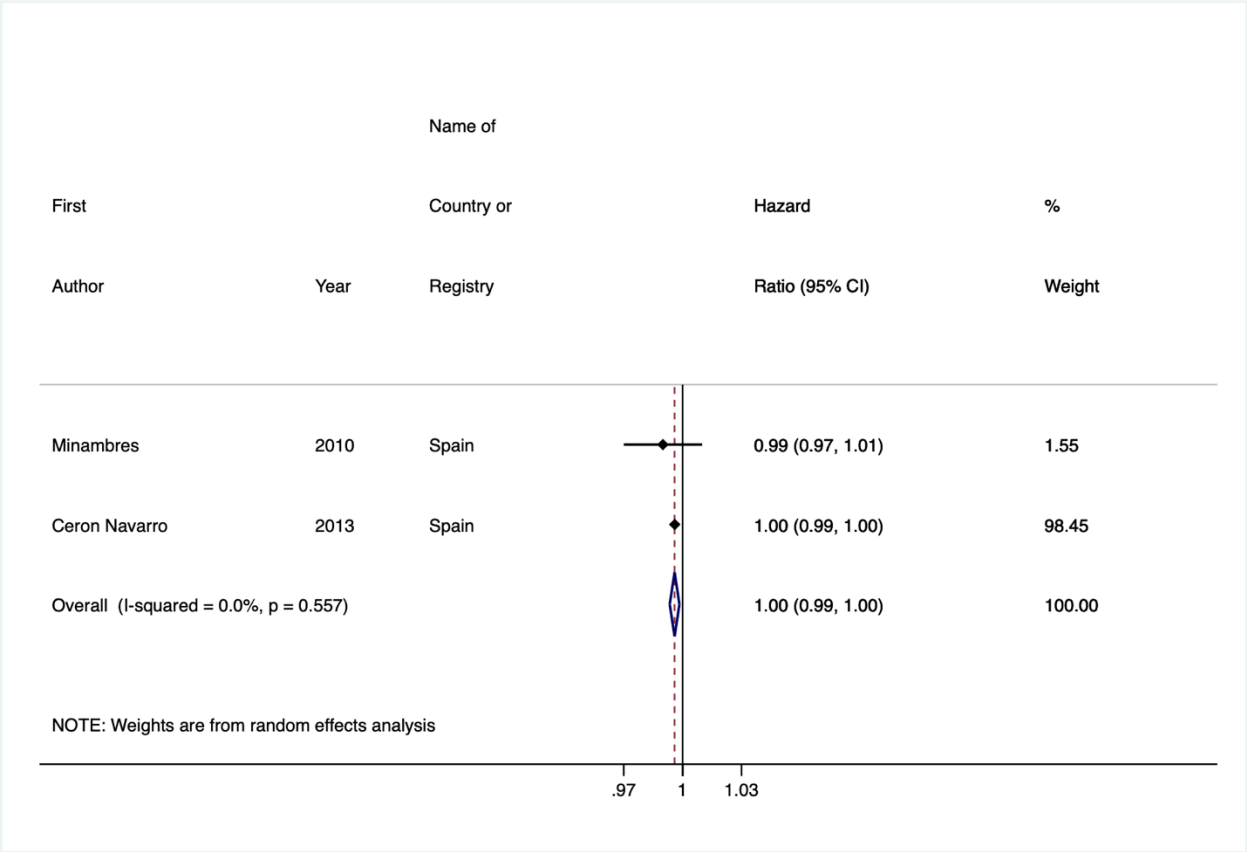


APPENDIX V – Meta-analysis of recipient pre-transplant systolic pulmonary artery pressure



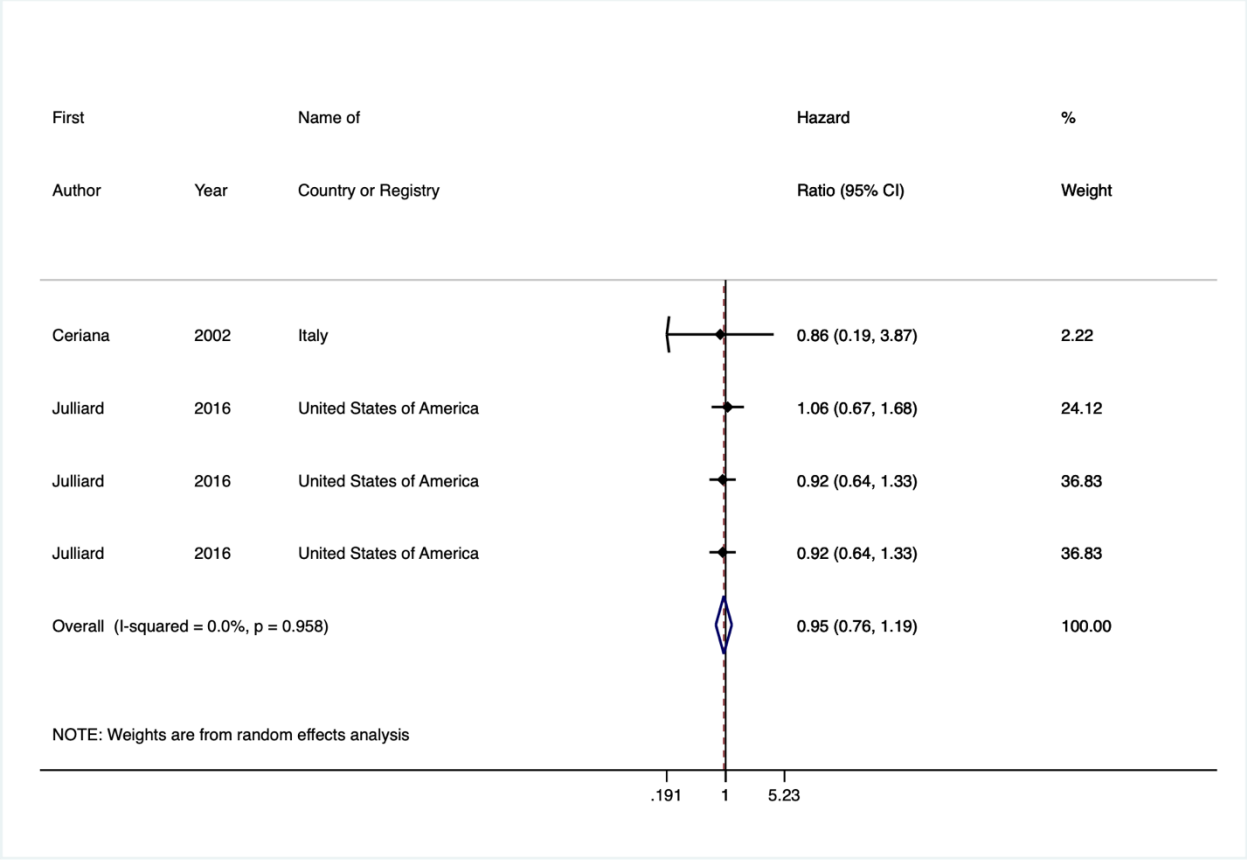
HR per 10 mmHg increase

APPENDIX W – Meta-analysis of recipient pre-transplant PaO_2 / FiO_2



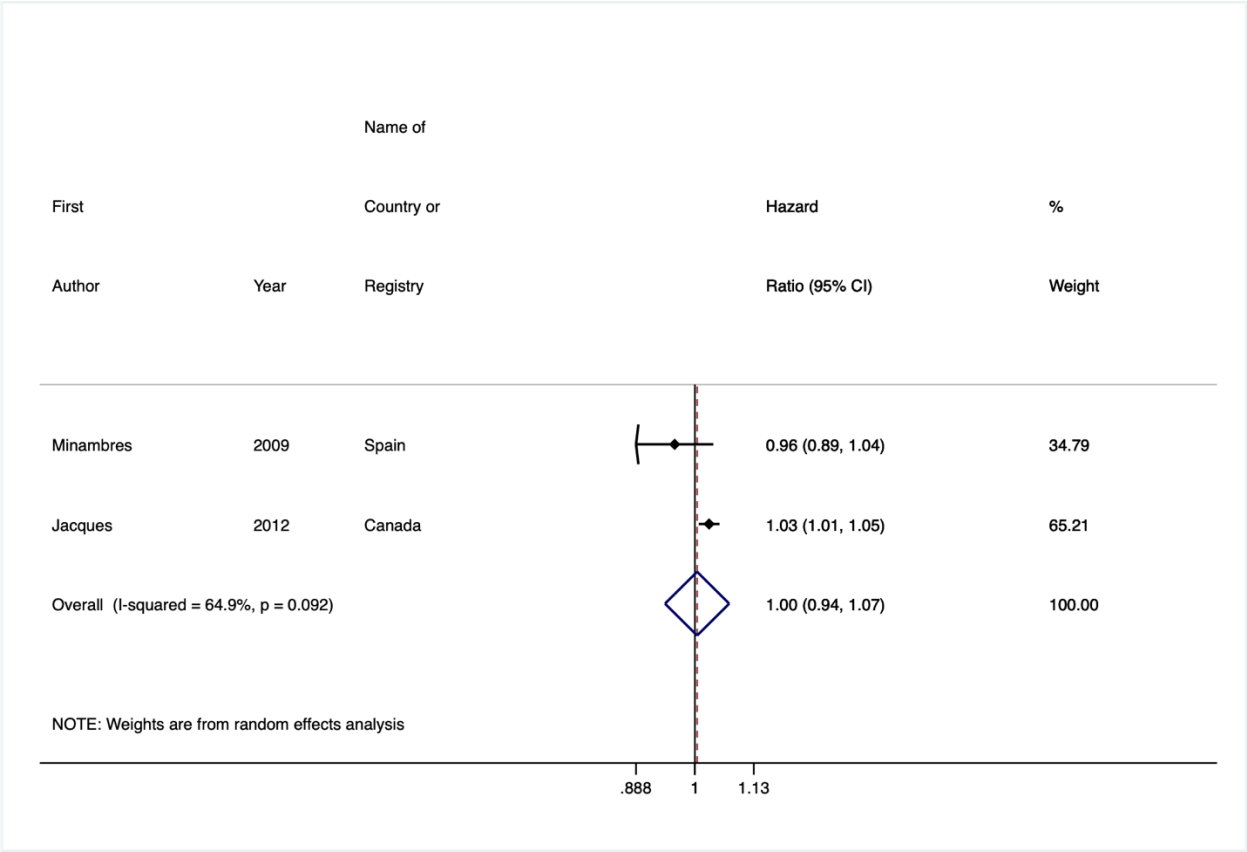
HR for 1 unit increase

APPENDIX X – Meta-Analysis of recipient pre-transplant mean pulmonary artery pressure



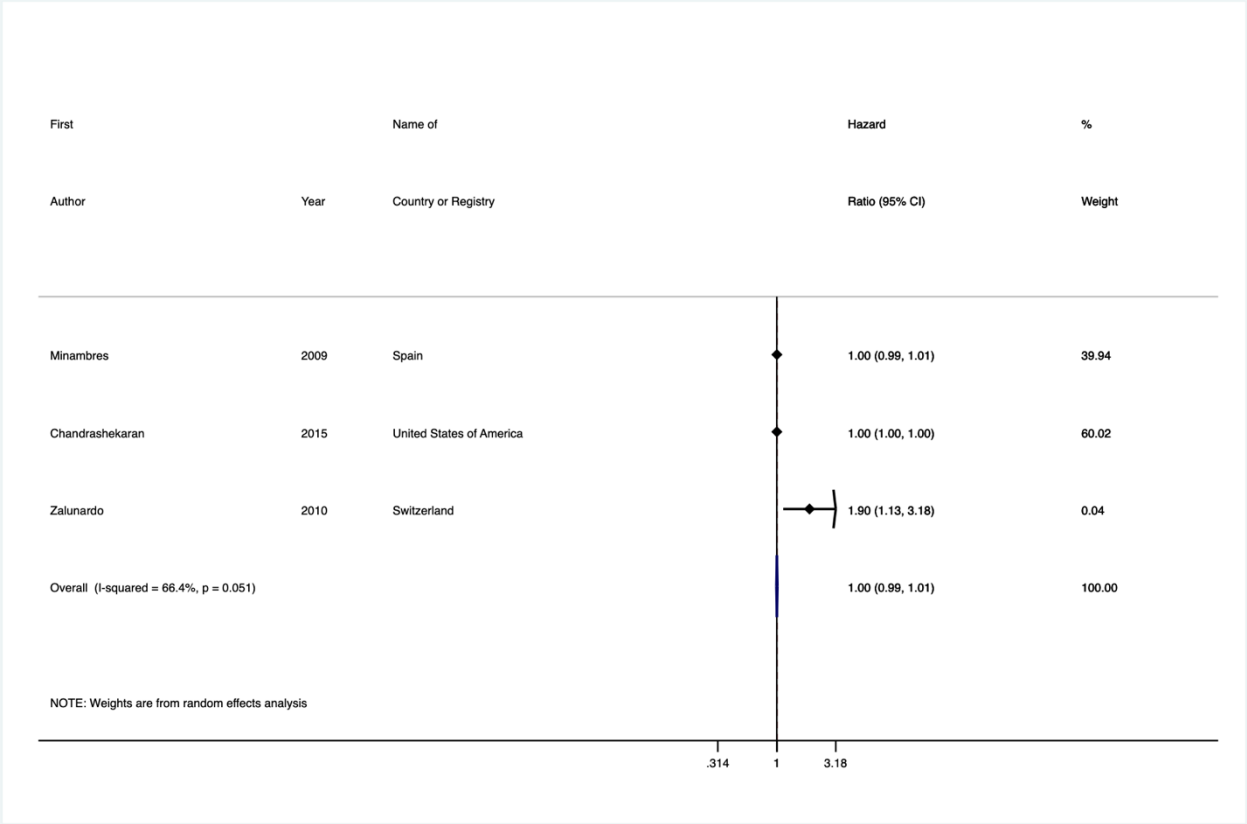
HR for >27 mmHg vs <27 mmHg

APPENDIX Y – Meta-analysis of recipient pre-transplant length of ICU stay



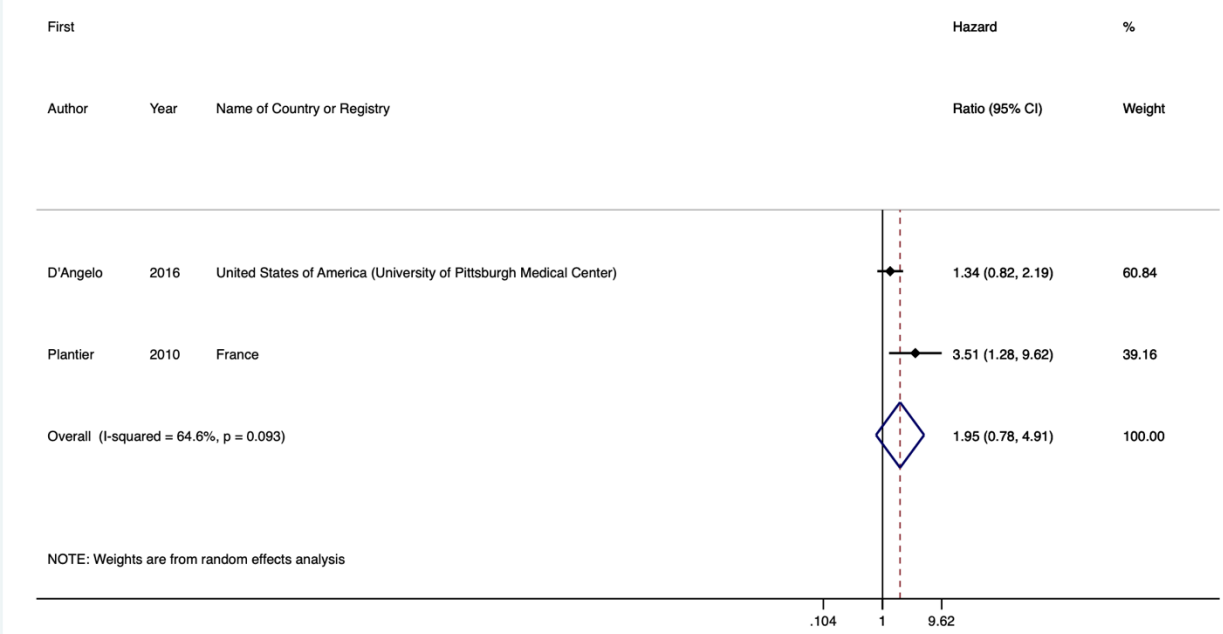
HR per 1-day increase

APPENDIX Z – Meta-analysis of recipient time on mechanical ventilation pre-transplant

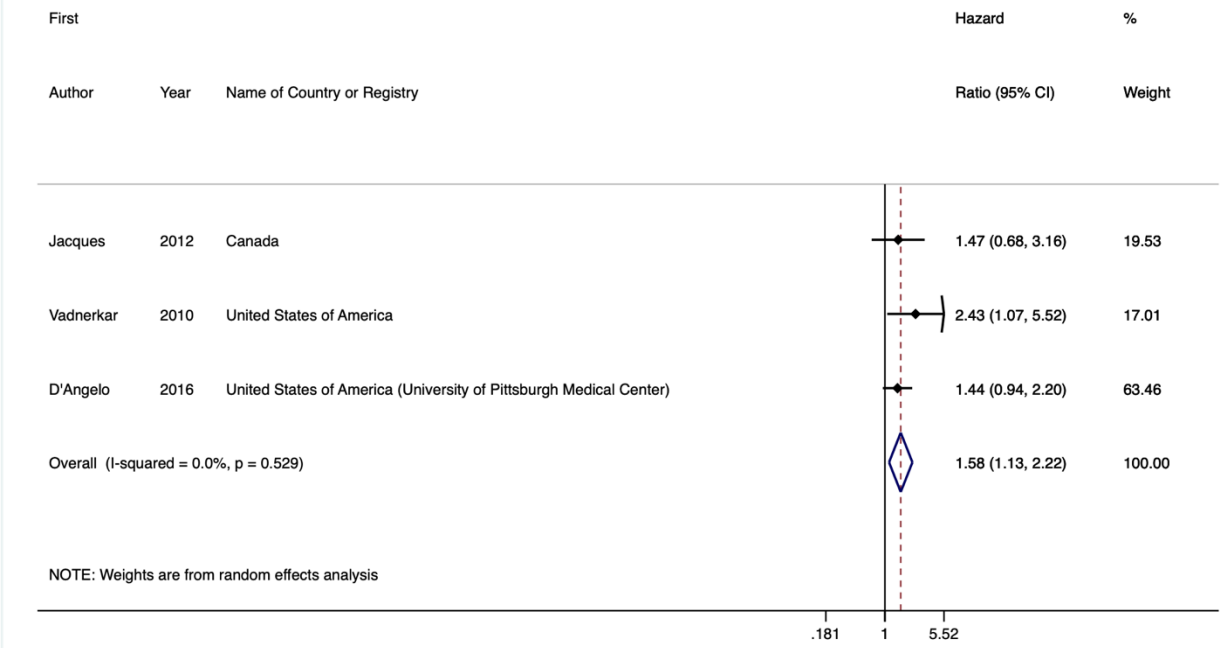


HR per 1 hour increase

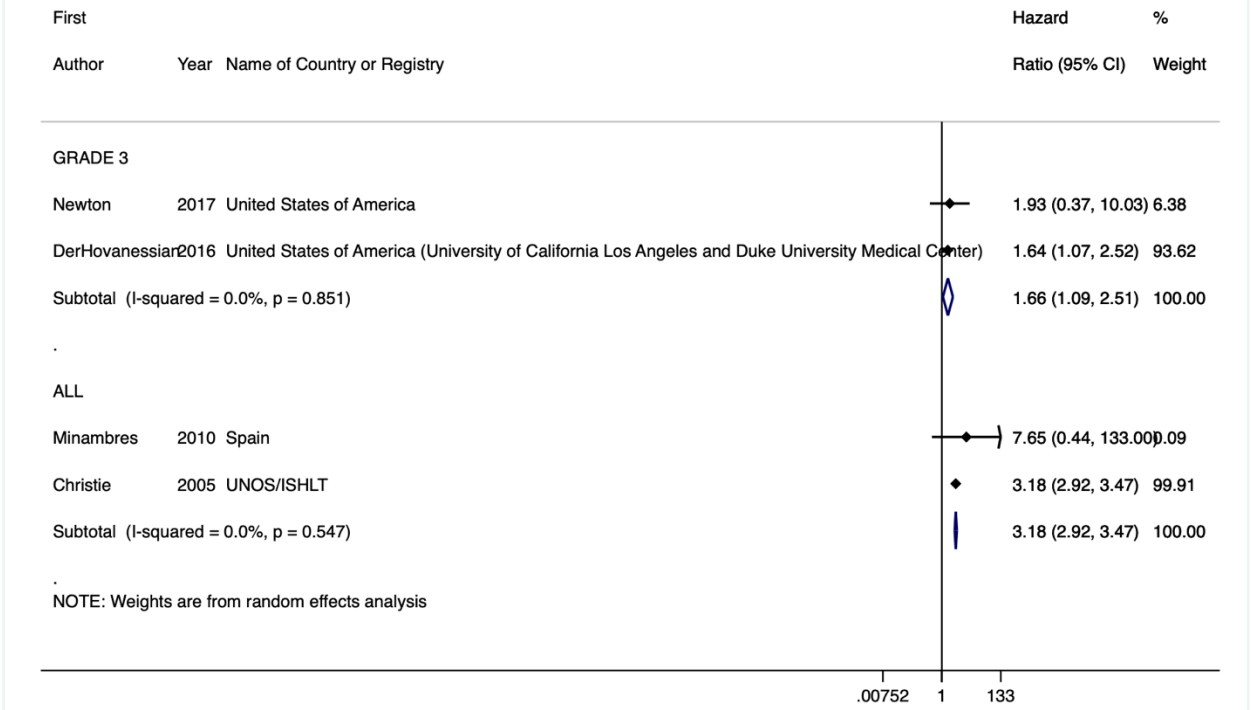
APPENDIX AA – Meta-analysis of recipient pre-transplant arrythmias



APPENDIX AB – Meta-analysis of recipient pre-transplant coronary artery disease

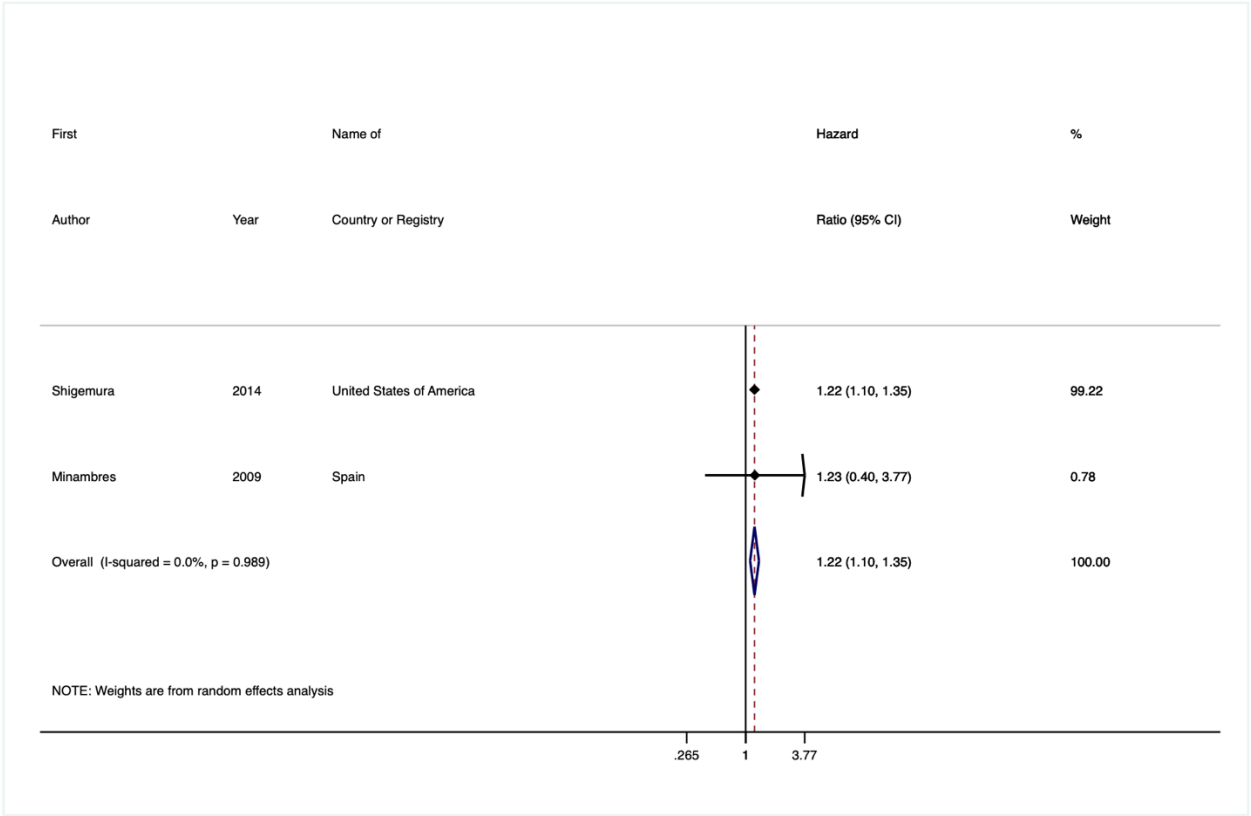


APPENDIX AC – Meta-analysis of post-transplant primary graft dysfunction

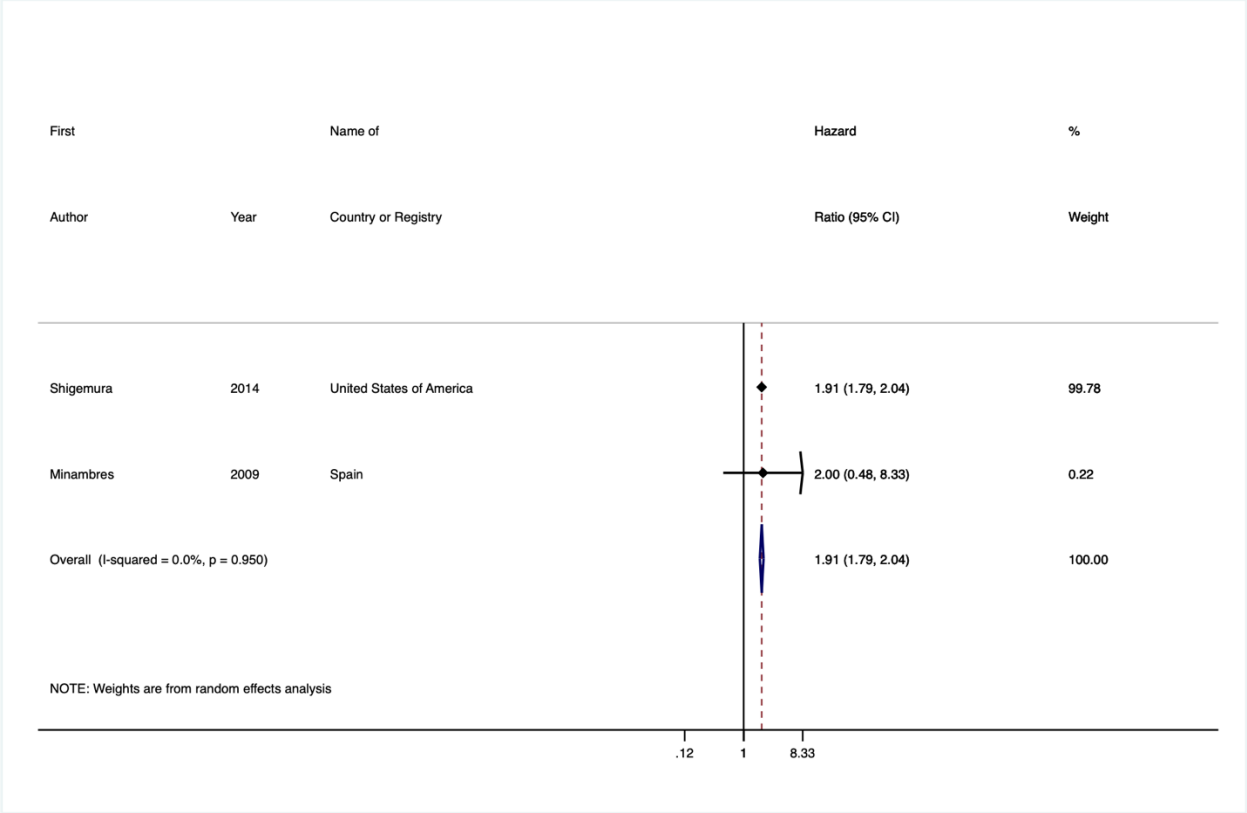


APPENDIX AD – Meta-analysis of post-transplant acute rejection

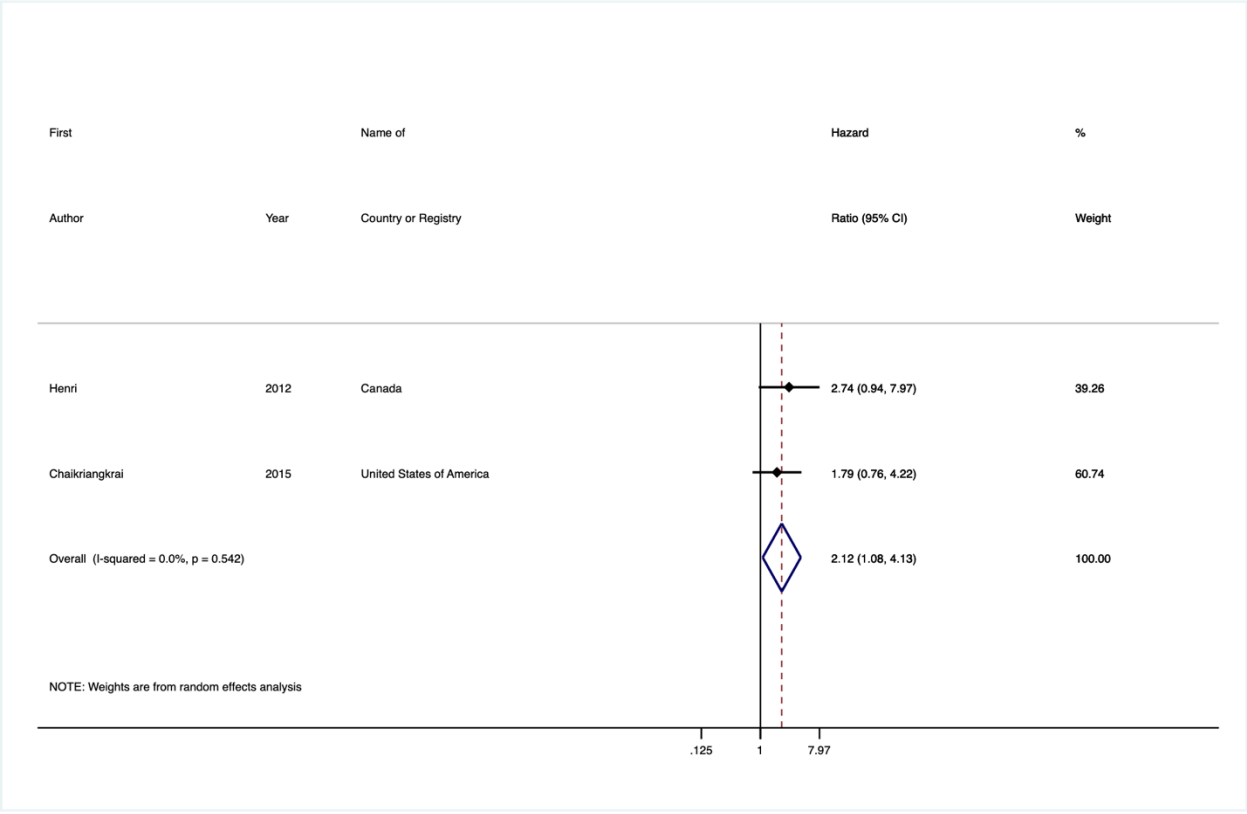
APPENDIX AE – Meta-analysis of post-transplant re-exploration for bleeding



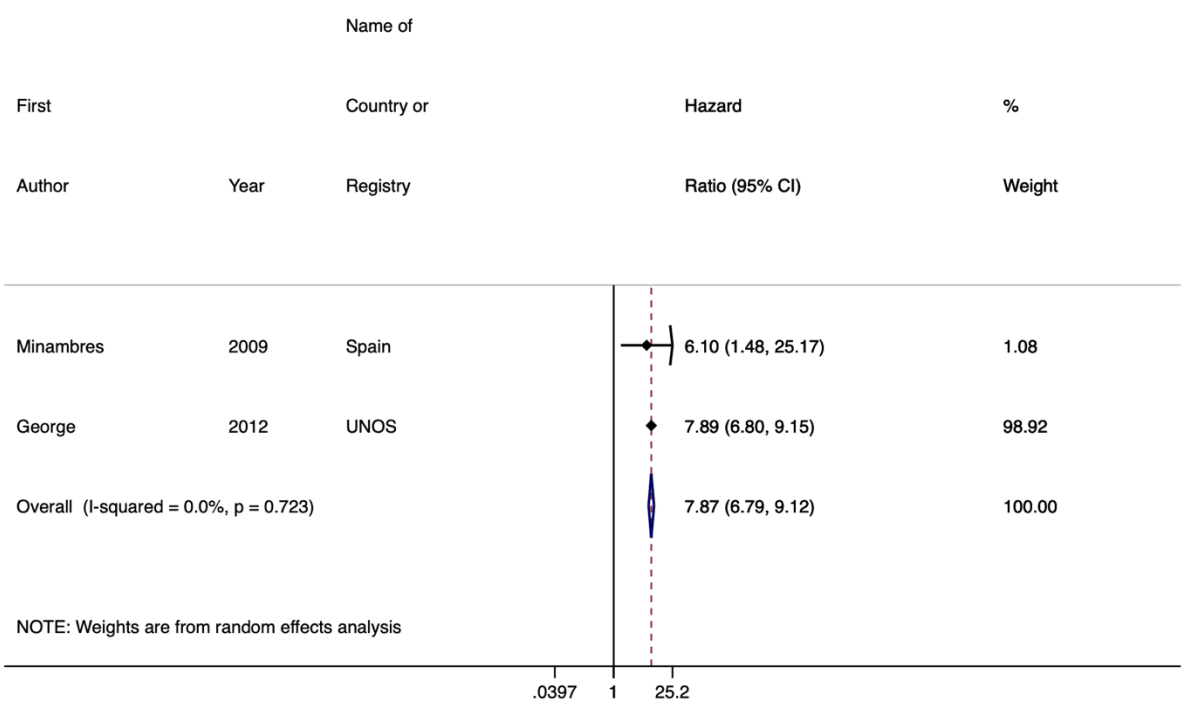
APPENDIX AF – Meta-analysis of post-transplant requirement for extra-corporeal membrane oxygenation



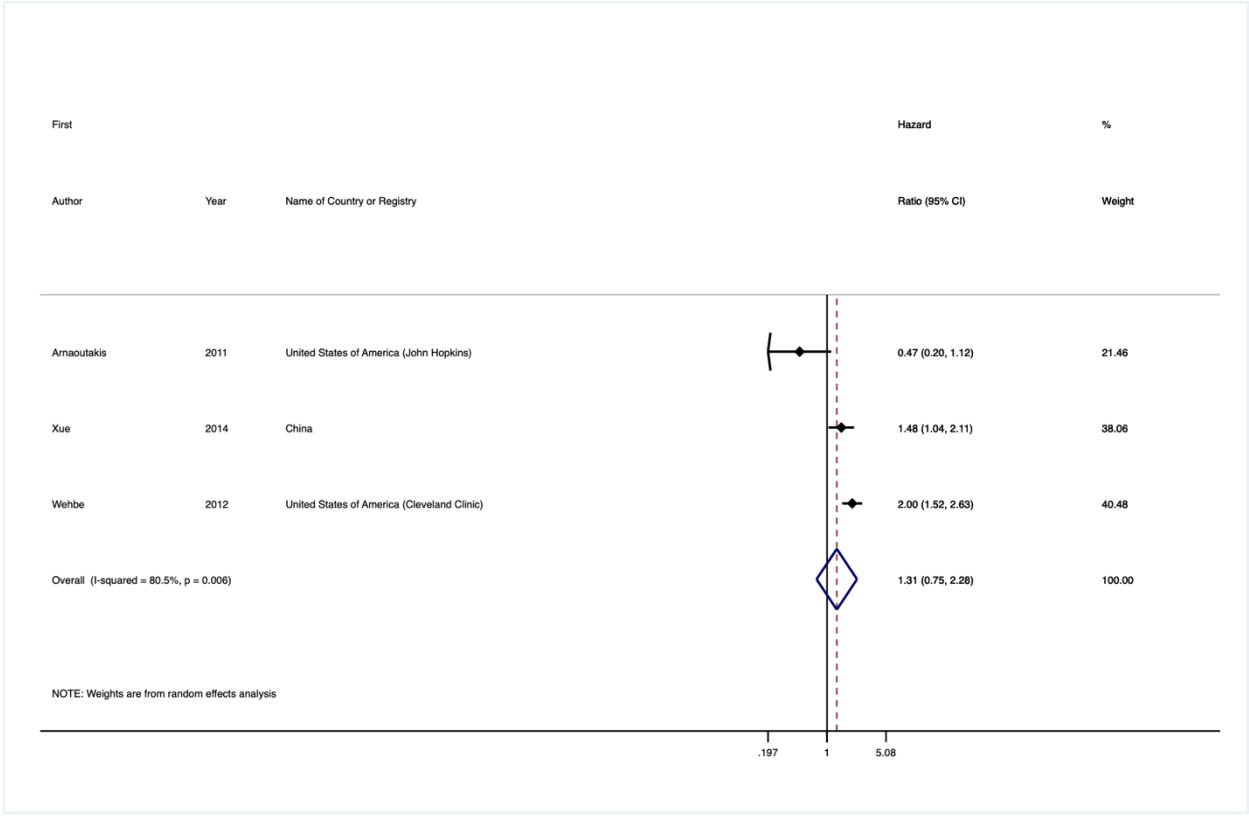
APPENDIX AG – Meta-analysis of post-transplant requirement for vasopressin



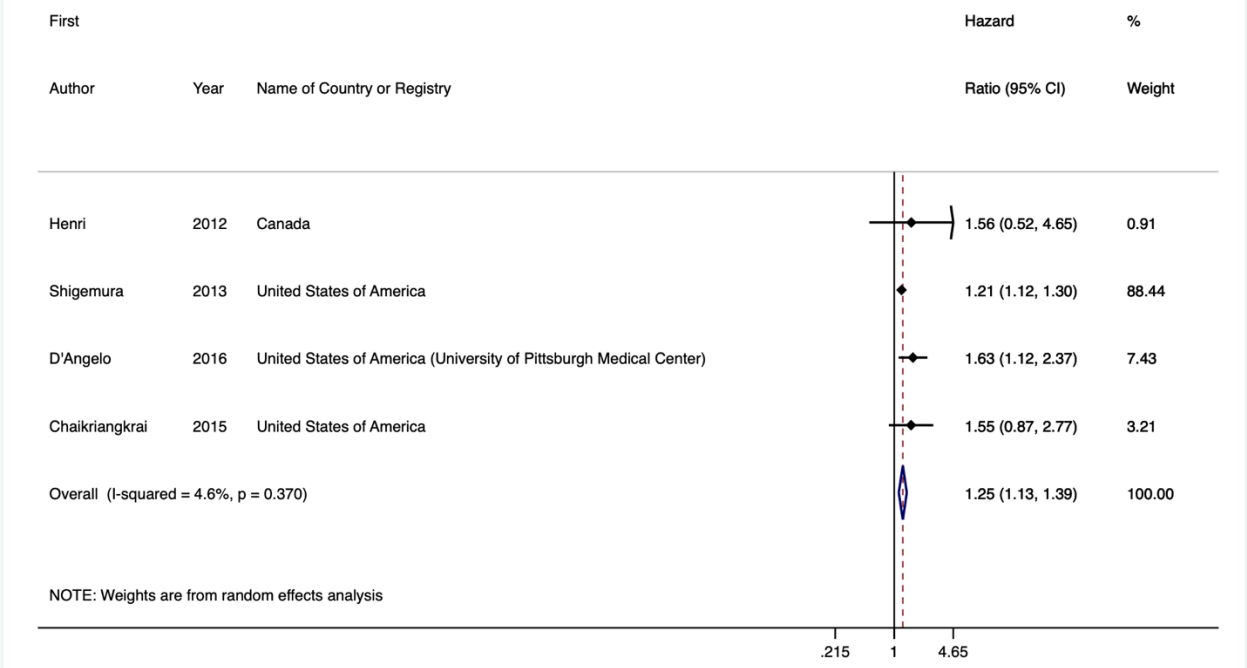
APPENDIX AH – Meta-analysis for post-transplant requirement for dialysis



APPENDIX AI – Meta-analysis for post-transplant acute kidney injury

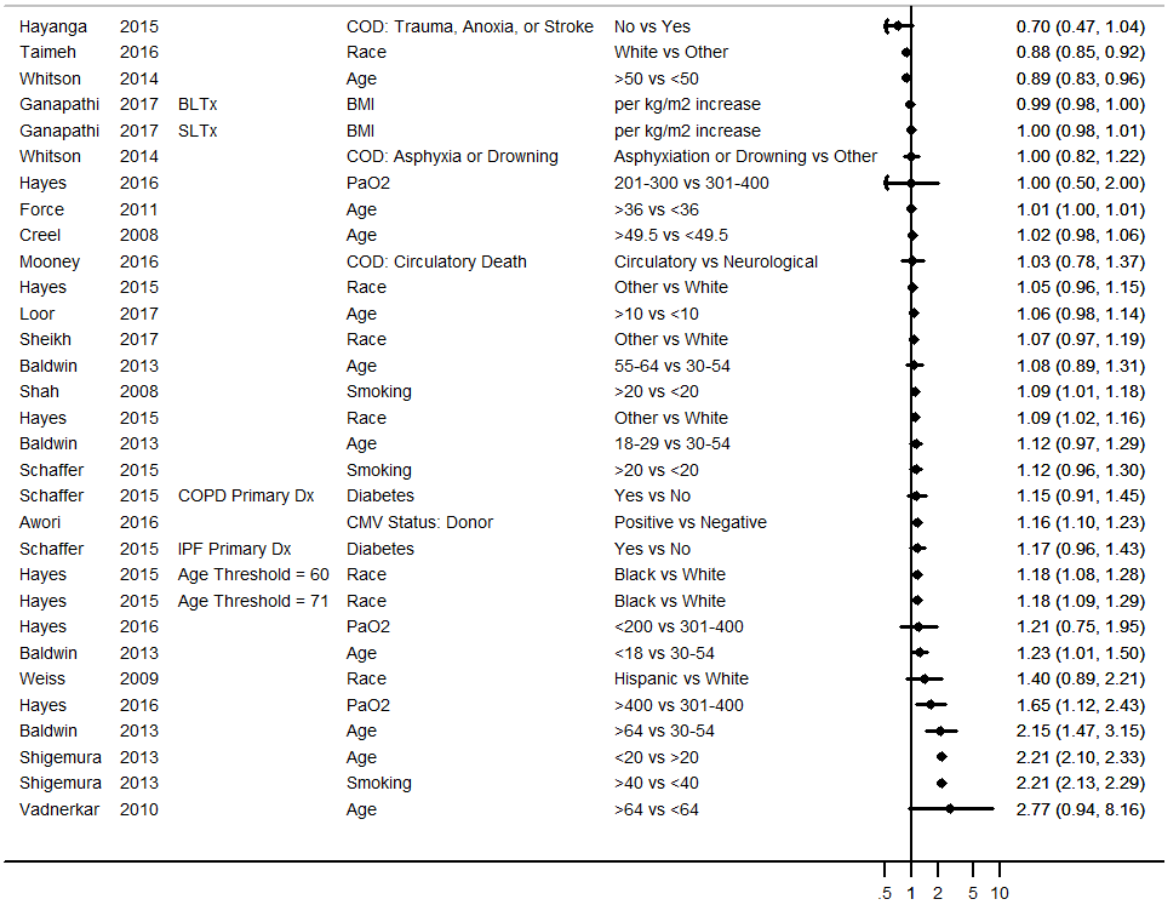


APPENDIX AJ – Meta-analysis for post-transplant cardiac complications



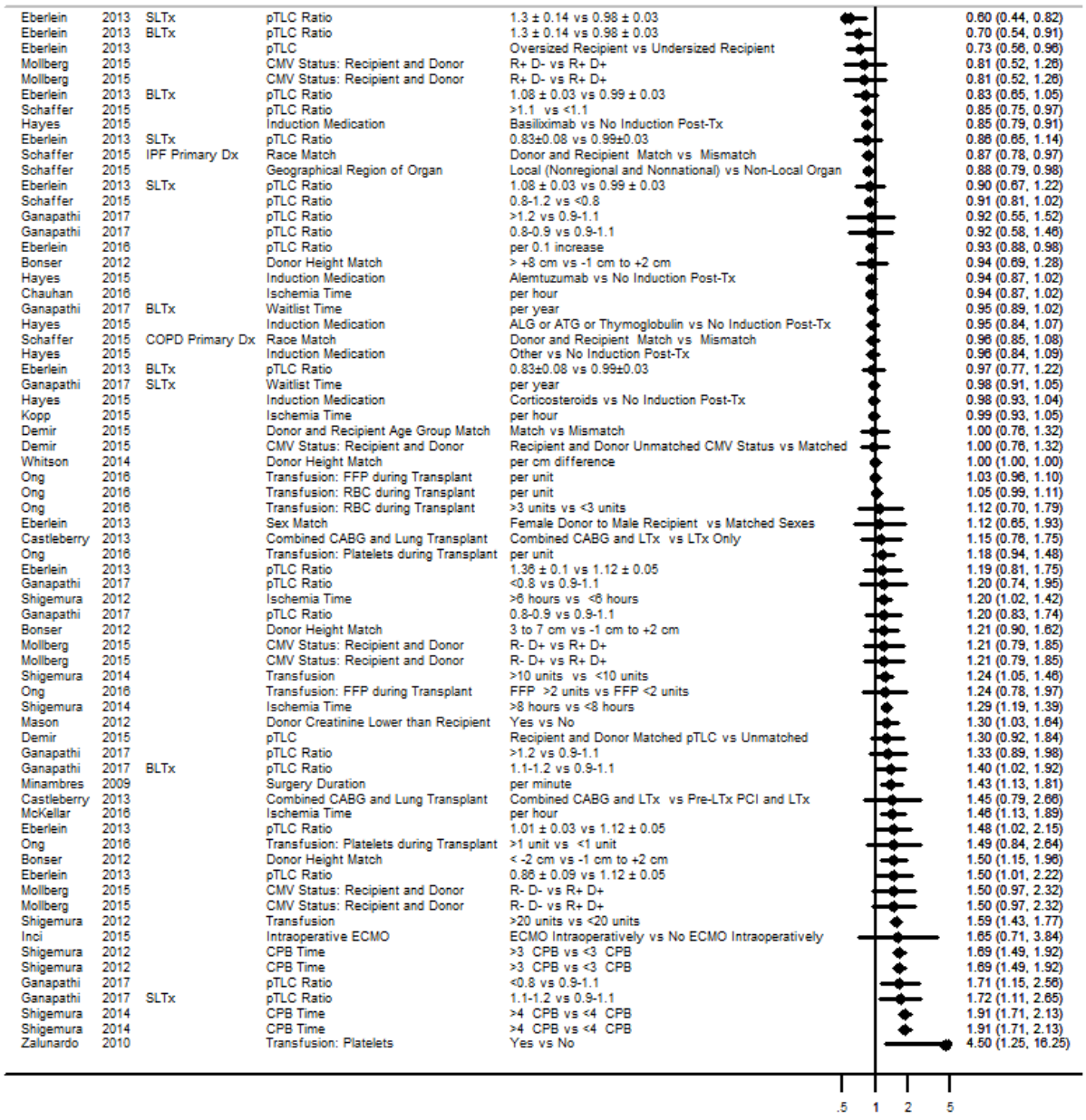
APPENDIX AK – List of unique predictors identified by one study related to donors

First Author	Year	Stratified Group	Predictor	Category	Hazard Ratio (95% CI)
Hayanga	2015		COD: Trauma, Anoxia, or Stroke	No vs Yes	0.70 (0.47, 1.04)
Taimeh	2016		Race	White vs Other	0.88 (0.85, 0.92)
Whitson	2014		Age	>50 vs <50	0.89 (0.83, 0.96)
Ganapathi	2017	BLTx	BMI	per kg/m2 increase	0.99 (0.98, 1.00)
Ganapathi	2017	SLTx	BMI	per kg/m2 increase	1.00 (0.98, 1.01)
Whitson	2014		COD: Asphyxia or Drowning	Asphyxiation or Drowning vs Other	1.00 (0.82, 1.22)
Hayes	2016		PaO2	201-300 vs 301-400	1.00 (0.50, 2.00)
Force	2011		Age	>36 vs <36	1.01 (1.00, 1.01)
Creel	2008		Age	>49.5 vs <49.5	1.02 (0.98, 1.06)
Mooney	2016		COD: Circulatory Death	Circulatory vs Neurological	1.03 (0.78, 1.37)
Hayes	2015		Race	Other vs White	1.05 (0.96, 1.15)
Loor	2017		Age	>10 vs <10	1.06 (0.98, 1.14)
Sheikh	2017		Race	Other vs White	1.07 (0.97, 1.19)
Baldwin	2013		Age	55-64 vs 30-54	1.08 (0.89, 1.31)
Shah	2008		Smoking	>20 vs <20	1.09 (1.01, 1.18)
Hayes	2015		Race	Other vs White	1.09 (1.02, 1.16)
Baldwin	2013		Age	18-29 vs 30-54	1.12 (0.97, 1.29)
Schaffer	2015		Smoking	>20 vs <20	1.12 (0.96, 1.30)
Schaffer	2015	COPD Primary Dx	Diabetes	Yes vs No	1.15 (0.91, 1.45)
Awori	2016		CMV Status: Donor	Positive vs Negative	1.16 (1.10, 1.23)
Schaffer	2015	IPF Primary Dx	Diabetes	Yes vs No	1.17 (0.96, 1.43)
Hayes	2015	Age Threshold = 60	Race	Black vs White	1.18 (1.08, 1.28)
Hayes	2015	Age Threshold = 71	Race	Black vs White	1.18 (1.09, 1.29)
Hayes	2016		PaO2	<200 vs 301-400	1.21 (0.75, 1.95)
Baldwin	2013		Age	<18 vs 30-54	1.23 (1.01, 1.50)
Weiss	2009		Race	Hispanic vs White	1.40 (0.89, 2.21)
Hayes	2016		PaO2	>400 vs 301-400	1.65 (1.12, 2.43)
Baldwin	2013		Age	>64 vs 30-54	2.15 (1.47, 3.15)
Shigemura	2013		Age	<20 vs >20	2.21 (2.10, 2.33)
Shigemura	2013		Smoking	>40 vs <40	2.21 (2.13, 2.29)
Vadnerkar	2010		Age	>64 vs <64	2.77 (0.94, 8.16)

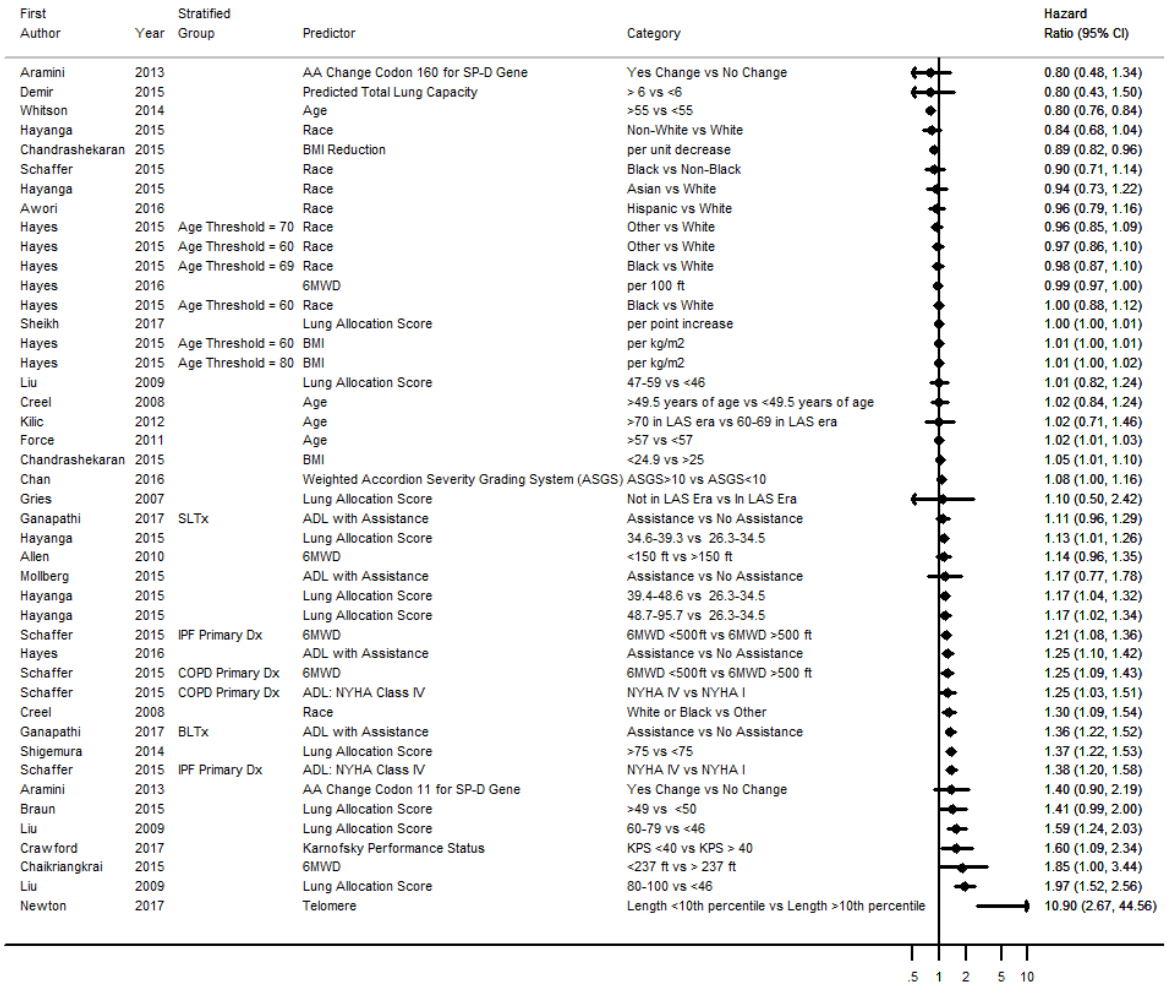


APPENDIX AL – List of unique predictors identified by one study related to transplantation

First Author	Year	Stratified Group	Predictor	Category	Hazard Ratio (95% CI)
Eberlein	2013	SLTx	pTLC Ratio	1.3 ± 0.14 vs 0.98 ± 0.03	0.60 (0.44, 0.82)
Eberlein	2013	BLTx	pTLC Ratio	1.3 ± 0.14 vs 0.98 ± 0.03	0.70 (0.54, 0.91)
Eberlein	2013		pTLC	Oversized Recipient vs Undersized Recipient	0.73 (0.56, 0.96)
Mollberg	2015		CMV Status: Recipient and Donor	R+ D- vs R+ D+	0.81 (0.52, 1.26)
Mollberg	2015		CMV Status: Recipient and Donor	R+ D- vs R+ D+	0.81 (0.52, 1.26)
Eberlein	2013	BLTx	pTLC Ratio	1.08 ± 0.03 vs 0.99 ± 0.03	0.83 (0.65, 1.05)
Schaffer	2015		pTLC Ratio	>1.1 vs <1.1	0.85 (0.75, 0.97)
Hayes	2015		Induction Medication	Basiliximab vs No Induction Post-Tx	0.85 (0.79, 0.91)
Eberlein	2013	SLTx	pTLC Ratio	0.83±0.08 vs 0.99±0.03	0.86 (0.65, 1.14)
Schaffer	2015	IPF Primary Dx	Race Match	Donor and Recipient Match vs Mismatch	0.87 (0.78, 0.97)
Schaffer	2015		Geographical Region of Organ	Local (Nonregional and Nonnational) vs Non-Local Organ	0.88 (0.79, 0.98)
Eberlein	2013	SLTx	pTLC Ratio	1.08 ± 0.03 vs 0.99 ± 0.03	0.90 (0.67, 1.22)
Schaffer	2015		pTLC Ratio	0.8-1.2 vs <0.8	0.91 (0.81, 1.02)
Ganapathi	2017		pTLC Ratio	>1.2 vs 0.9-1.1	0.92 (0.55, 1.52)
Ganapathi	2017		pTLC Ratio	0.8-0.9 vs 0.9-1.1	0.92 (0.58, 1.46)
Eberlein	2016		pTLC Ratio	per 0.1 increase	0.93 (0.88, 0.98)
Bonser	2012		Donor Height Match	> +8 cm vs -1 cm to +2 cm	0.94 (0.69, 1.28)
Hayes	2015		Induction Medication	Alemtuzumab vs No Induction Post-Tx	0.94 (0.87, 1.02)
Chauhan	2016		Ischemia Time	per hour	0.94 (0.87, 1.02)
Ganapathi	2017	BLTx	Waitlist Time	per year	0.95 (0.89, 1.02)
Hayes	2015		Induction Medication	ALG or ATG or Thymoglobulin vs No Induction Post-Tx	0.95 (0.84, 1.07)
Schaffer	2015	COPD Primary Dx	Race Match	Donor and Recipient Match vs Mismatch	0.96 (0.85, 1.08)
Hayes	2015		Induction Medication	Other vs No Induction Post-Tx	0.96 (0.84, 1.09)
Eberlein	2013	BLTx	pTLC Ratio	0.83±0.08 vs 0.99±0.03	0.97 (0.77, 1.22)
Ganapathi	2017	SLTx	Waitlist Time	per year	0.98 (0.91, 1.05)
Hayes	2015		Induction Medication	Corticosteroids vs No Induction Post-Tx	0.98 (0.93, 1.04)
Kopp	2015		Ischemia Time	per hour	0.99 (0.93, 1.05)
Demir	2015		Donor and Recipient Age Group Match	Match vs Mismatch	1.00 (0.76, 1.32)
Demir	2015		CMV Status: Recipient and Donor	Recipient and Donor Unmatched CMV Status vs Matched	1.00 (0.76, 1.32)
Whitson	2014		Donor Height Match	per cm difference	1.00 (1.00, 1.00)
Ong	2016		Transfusion: FFP during Transplant	per unit	1.03 (0.96, 1.10)
Ong	2016		Transfusion: RBC during Transplant	per unit	1.05 (0.99, 1.11)
Ong	2016		Transfusion: RBC during Transplant	>3 units vs <3 units	1.12 (0.70, 1.79)
Eberlein	2013		Sex Match	Female Donor to Male Recipient vs Matched Sexes	1.12 (0.65, 1.93)
Castleberry	2013		Combined CABG and Lung Transplant	Combined CABG and LTx vs LTx Only	1.15 (0.78, 1.75)
Ong	2016		Transfusion: Platelets during Transplant	per unit	1.18 (0.94, 1.48)
Eberlein	2013		pTLC Ratio	1.36 ± 0.1 vs 1.12 ± 0.05	1.19 (0.81, 1.75)
Ganapathi	2017		pTLC Ratio	<0.8 vs 0.9-1.1	1.20 (0.74, 1.95)
Shigemura	2012		Ischemia Time	>6 hours vs <6 hours	1.20 (1.02, 1.42)
Ganapathi	2017		pTLC Ratio	0.8-0.9 vs 0.9-1.1	1.20 (0.83, 1.74)
Bonser	2012		Donor Height Match	3 to 7 cm vs -1 cm to +2 cm	1.21 (0.90, 1.62)
Mollberg	2015		CMV Status: Recipient and Donor	R- D+ vs R+ D+	1.21 (0.79, 1.85)
Mollberg	2015		CMV Status: Recipient and Donor	R- D+ vs R+ D+	1.21 (0.79, 1.85)
Shigemura	2014		Transfusion	>10 units vs <10 units	1.24 (1.05, 1.46)
Ong	2016		Transfusion: FFP during Transplant	FFP >2 units vs FFP <2 units	1.24 (0.78, 1.97)
Shigemura	2014		Ischemia Time	>8 hours vs <8 hours	1.29 (1.19, 1.39)
Mason	2012		Donor Creatinine Lower than Recipient	Yes vs No	1.30 (1.03, 1.64)
Demir	2015		pTLC	Recipient and Donor Matched pTLC vs Unmatched	1.30 (0.92, 1.84)
Ganapathi	2017		pTLC Ratio	>1.2 vs 0.9-1.1	1.33 (0.89, 1.98)
Ganapathi	2017	BLTx	pTLC Ratio	1.1-1.2 vs 0.9-1.1	1.40 (1.02, 1.92)
Minambres	2009		Surgery Duration	per minute	1.43 (1.13, 1.81)
Castleberry	2013		Combined CABG and Lung Transplant	Combined CABG and LTx vs Pre-LTx PCI and LTx	1.45 (0.79, 2.66)
McKellar	2016		Ischemia Time	per hour	1.46 (1.13, 1.89)
Eberlein	2013		pTLC Ratio	1.01 ± 0.03 vs 1.12 ± 0.05	1.48 (1.02, 2.15)
Ong	2016		Transfusion: Platelets during Transplant	>1 unit vs <1 unit	1.49 (0.84, 2.64)
Bonser	2012		Donor Height Match	< -2 cm vs -1 cm to +2 cm	1.50 (1.15, 1.96)
Eberlein	2013		pTLC Ratio	0.86 ± 0.09 vs 1.12 ± 0.05	1.50 (1.01, 2.22)
Mollberg	2015		CMV Status: Recipient and Donor	R- D- vs R+ D+	1.50 (0.97, 2.32)
Mollberg	2015		CMV Status: Recipient and Donor	R- D- vs R+ D+	1.50 (0.97, 2.32)
Shigemura	2012		Transfusion	>20 units vs <20 units	1.59 (1.43, 1.77)
Inci	2015		Intraoperative ECMO	ECMO Intraoperatively vs No ECMO Intraoperatively	1.65 (0.71, 3.84)
Shigemura	2012		CPB Time	>3 CPB vs <3 CPB	1.69 (1.49, 1.92)
Shigemura	2012		CPB Time	>3 CPB vs <3 CPB	1.69 (1.49, 1.92)
Ganapathi	2017		pTLC Ratio	<0.8 vs 0.9-1.1	1.71 (1.15, 2.56)
Ganapathi	2017	SLTx	pTLC Ratio	1.1-1.2 vs 0.9-1.1	1.72 (1.11, 2.65)
Shigemura	2014		CPB Time	>4 CPB vs <4 CPB	1.91 (1.71, 2.13)
Shigemura	2014		CPB Time	>4 CPB vs <4 CPB	1.91 (1.71, 2.13)
Zalunardo	2010		Transfusion: Platelets	Yes vs No	4.50 (1.25, 16.25)

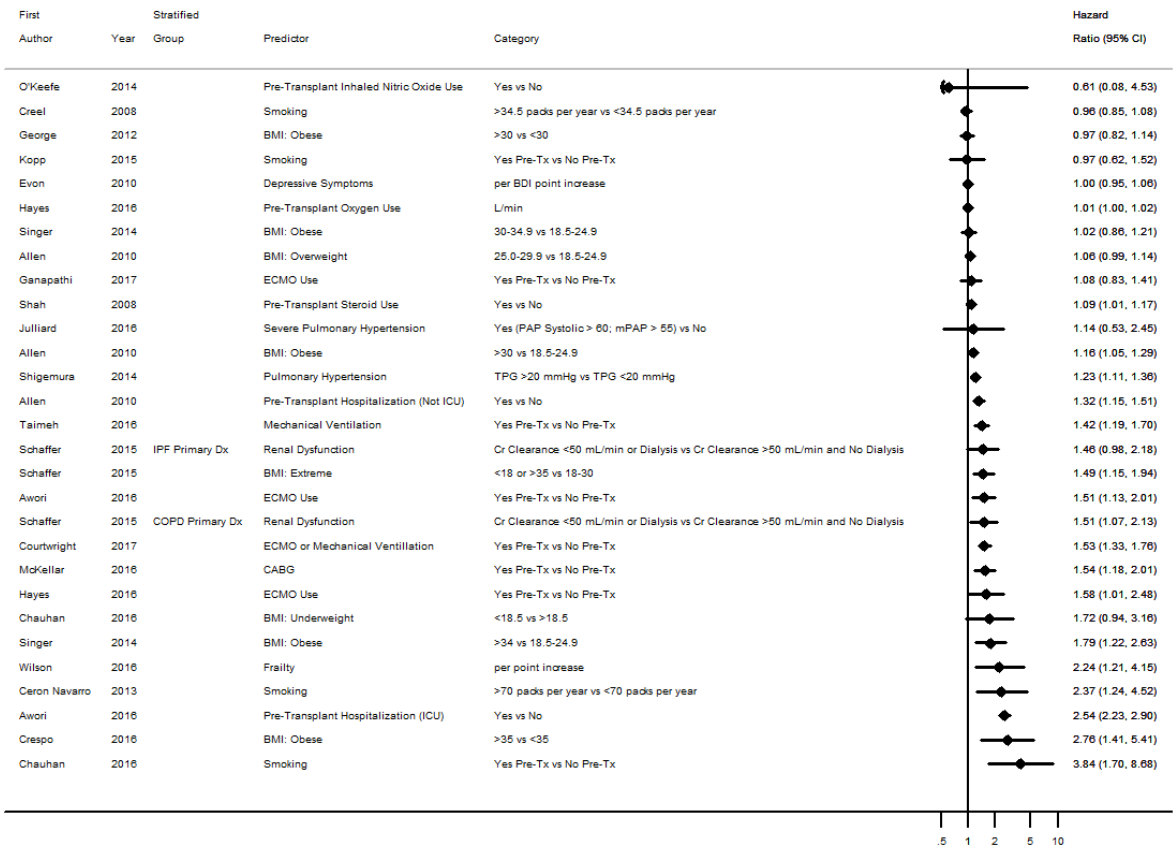


APPENDIX AM – List of unique predictors identified by one study related to recipient demographics



First Author	Year	Stratified Group	Predictor	Category	Hazard Ratio (95% CI)
Aramini	2013		AA Change Codon 160 for SP-D Gene	Yes Change vs No Change	0.80 (0.48, 1.34)
Demir	2015		Predicted Total Lung Capacity	> 6 vs <6	0.80 (0.43, 1.50)
Whitson	2014		Age	>55 vs <55	0.80 (0.76, 0.84)
Hayanga	2015		Race	Non-White vs White	0.84 (0.68, 1.04)
Chandrashekaran	2015		BMI Reduction	per unit decrease	0.89 (0.82, 0.96)
Schaffer	2015		Race	Black vs Non-Black	0.90 (0.71, 1.14)
Hayanga	2015		Race	Asian vs White	0.94 (0.73, 1.22)
Awori	2016		Race	Hispanic vs White	0.96 (0.79, 1.16)
Hayes	2015	Age Threshold = 70	Race	Other vs White	0.96 (0.85, 1.09)
Hayes	2015	Age Threshold = 60	Race	Other vs White	0.97 (0.86, 1.10)
Hayes	2015	Age Threshold = 69	Race	Black vs White	0.98 (0.87, 1.10)
Hayes	2016		6MWD	per 100 ft	0.99 (0.97, 1.00)
Hayes	2015	Age Threshold = 60	Race	Black vs White	1.00 (0.88, 1.12)
Sheikh	2017		Lung Allocation Score	per point increase	1.00 (1.00, 1.01)
Hayes	2015	Age Threshold = 60	BMI	per kg/m2	1.01 (1.00, 1.01)
Hayes	2015	Age Threshold = 80	BMI	per kg/m2	1.01 (1.00, 1.02)
Liu	2009		Lung Allocation Score	47-59 vs <46	1.01 (0.82, 1.24)
Creel	2008		Age	>49.5 years of age vs <49.5 years of age	1.02 (0.84, 1.24)
Kilic	2012		Age	>70 in LAS era vs 60-69 in LAS era	1.02 (0.71, 1.46)
Force	2011		Age	>57 vs <57	1.02 (1.01, 1.03)
Chandrashekaran	2015		BMI	<24.9 vs >25	1.05 (1.01, 1.10)
Chan	2016		Weighted Accordion Severity Grading System (ASGS)	ASGS>10 vs ASGS<10	1.08 (1.00, 1.16)
Gries	2007		Lung Allocation Score	Not in LAS Era vs In LAS Era	1.10 (0.50, 2.42)
Ganapathi	2017	SLTx	ADL with Assistance	Assistance vs No Assistance	1.11 (0.96, 1.29)
Hayanga	2015		Lung Allocation Score	34.6-39.3 vs 26.3-34.5	1.13 (1.01, 1.26)
Allen	2010		6MWD	<150 ft vs >150 ft	1.14 (0.96, 1.35)
Mollberg	2015		ADL with Assistance	Assistance vs No Assistance	1.17 (0.77, 1.78)
Hayanga	2015		Lung Allocation Score	39.4-48.6 vs 26.3-34.5	1.17 (1.04, 1.32)
Hayanga	2015		Lung Allocation Score	48.7-95.7 vs 26.3-34.5	1.17 (1.02, 1.34)
Schaffer	2015	IPF Primary Dx	6MWD	6MWD <500ft vs 6MWD >500 ft	1.21 (1.08, 1.36)
Hayes	2016		ADL with Assistance	Assistance vs No Assistance	1.25 (1.10, 1.42)
Schaffer	2015	COPD Primary Dx	6MWD	6MWD <500ft vs 6MWD >500 ft	1.25 (1.09, 1.43)
Schaffer	2015	COPD Primary Dx	ADL: NYHA Class IV	NYHA IV vs NYHA I	1.25 (1.03, 1.51)
Creel	2008		Race	White or Black vs Other	1.30 (1.09, 1.54)
Ganapathi	2017	BLTx	ADL with Assistance	Assistance vs No Assistance	1.36 (1.22, 1.52)
Shigemura	2014		Lung Allocation Score	>75 vs <75	1.37 (1.22, 1.53)
Schaffer	2015	IPF Primary Dx	ADL: NYHA Class IV	NYHA IV vs NYHA I	1.38 (1.20, 1.58)
Aramini	2013		AA Change Codon 11 for SP-D Gene	Yes Change vs No Change	1.40 (0.90, 2.19)
Braun	2015		Lung Allocation Score	>49 vs <50	1.41 (0.99, 2.00)
Liu	2009		Lung Allocation Score	60-79 vs <46	1.59 (1.24, 2.03)
Crawford	2017		Karnofsky Performance Status	KPS <40 vs KPS > 40	1.60 (1.09, 2.34)
Chaikriangkrai	2015		6MWD	<237 ft vs > 237 ft	1.85 (1.00, 3.44)
Liu	2009		Lung Allocation Score	80-100 vs <46	1.97 (1.52, 2.56)
Newton	2017		Telomere	Length <10th percentile vs Length >10th percentile	10.90 (2.67, 44.56)

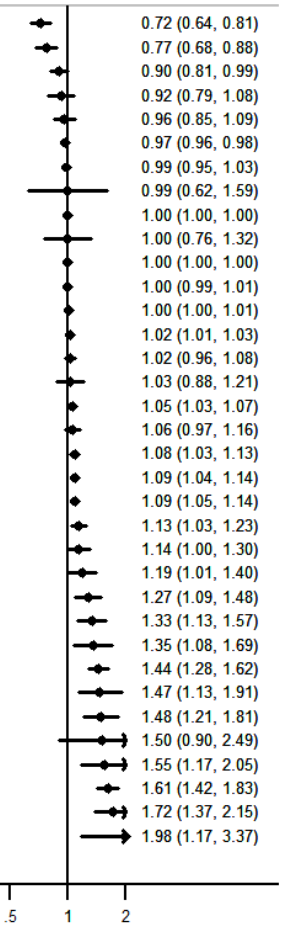
APPENDIX AN – List of unique predictors identified by one study related to recipient comorbidities



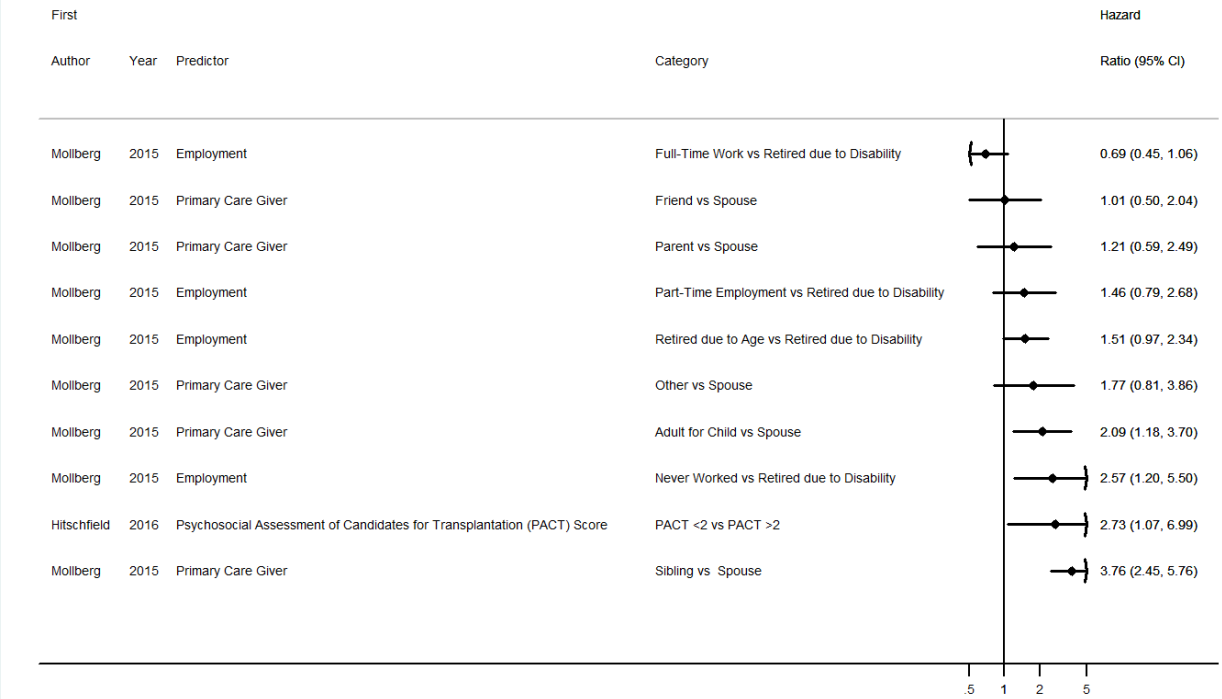
First Author	Year	Stratified Group	Predictor	Category	Hazard Ratio (95% CI)
O'Keefe	2014		Pre-Transplant Inhaled Nitric Oxide Use	Yes vs No	0.61 (0.08, 4.53)
Creel	2008		Smoking	>34.5 packs per year vs <34.5 packs per year	0.96 (0.85, 1.08)
George	2012		BMI: Obese	>30 vs <30	0.97 (0.82, 1.14)
Kopp	2015		Smoking	Yes Pre-Tx vs No Pre-Tx	0.97 (0.62, 1.52)
Evon	2010		Depressive Symptoms	per BDI point increase	1.00 (0.95, 1.06)
Hayes	2016		Pre-Transplant Oxygen Use	L/min	1.01 (1.00, 1.02)
Singer	2014		BMI: Obese	30-34.9 vs 18.5-24.9	1.02 (0.86, 1.21)
Allen	2010		BMI: Overweight	25.0-29.9 vs 18.5-24.9	1.06 (0.99, 1.14)
Ganapathi	2017		ECMO Use	Yes Pre-Tx vs No Pre-Tx	1.08 (0.83, 1.41)
Shah	2008		Pre-Transplant Steroid Use	Yes vs No	1.09 (1.01, 1.17)
Julliard	2016		Severe Pulmonary Hypertension	Yes (PAP Systolic > 60; mPAP > 55) vs No	1.14 (0.53, 2.45)
Allen	2010		BMI: Obese	>30 vs 18.5-24.9	1.16 (1.05, 1.29)
Shigemura	2014		Pulmonary Hypertension	TPG >20 mmHg vs TPG <20 mmHg	1.23 (1.11, 1.36)
Allen	2010		Pre-Transplant Hospitalization (Not ICU)	Yes vs No	1.32 (1.15, 1.51)
Taimeh	2016		Mechanical Ventilation	Yes Pre-Tx vs No Pre-Tx	1.42 (1.19, 1.70)
Schaffer	2015	IPF Primary Dx	Renal Dysfunction	Cr Clearance <50 mL/min or Dialysis vs Cr Clearance >50 mL/min and No Dialysis	1.46 (0.98, 2.18)
Schaffer	2015		BMI: Extreme	<18 or >35 vs 18-30	1.49 (1.15, 1.94)
Awori	2016		ECMO Use	Yes Pre-Tx vs No Pre-Tx	1.51 (1.13, 2.01)
Schaffer	2015	COPD Primary Dx	Renal Dysfunction	Cr Clearance <50 mL/min or Dialysis vs Cr Clearance >50 mL/min and No Dialysis	1.51 (1.07, 2.13)
Courtwright	2017		ECMO or Mechanical Ventillation	Yes Pre-Tx vs No Pre-Tx	1.53 (1.33, 1.76)
McKellar	2016		CABG	Yes Pre-Tx vs No Pre-Tx	1.54 (1.18, 2.01)
Hayes	2016		ECMO Use	Yes Pre-Tx vs No Pre-Tx	1.58 (1.01, 2.48)
Chauhan	2016		BMI: Underweight	<18.5 vs >18.5	1.72 (0.94, 3.16)
Singer	2014		BMI: Obese	>34 vs 18.5-24.9	1.79 (1.22, 2.63)
Wilson	2016		Frailty	per point increase	2.24 (1.21, 4.15)
Ceron Navarro	2013		Smoking	>70 packs per year vs <70 packs per year	2.37 (1.24, 4.52)
Awori	2016		Pre-Transplant Hospitalization (ICU)	Yes vs No	2.54 (2.23, 2.90)
Crespo	2016		BMI: Obese	>35 vs <35	2.76 (1.41, 5.41)
Chauhan	2016		Smoking	Yes Pre-Tx vs No Pre-Tx	3.84 (1.70, 8.68)

APPENDIX AO – List of unique predictors identified by one study related to recipient physiology

First Author	Year	Stratified Group	Predictor	Category	Hazard Ratio (95% CI)
Hayanga	2015		eGFR	>90 vs <90	0.72 (0.64, 0.81)
Hayanga	2015		eGFR	60-90 vs <60	0.77 (0.68, 0.88)
Demir	2015		ABO Status	Blood Group A vs Other Blood Group	0.90 (0.81, 0.99)
Hayes	2015		Pre-Transplant mPAP	<25 mmHg vs >25 mmHg	0.92 (0.79, 1.08)
Schaffer	2015		Pre-Transplant mPAP	<30 mmHg vs >30 mmHg	0.96 (0.85, 1.09)
Atkins	2010		Pre-Transplant FEV1	L/sec	0.97 (0.96, 0.98)
Hayes	2016		Pre-Transplant Cardiac Output	L/min	0.99 (0.95, 1.03)
Hayes	2016		PRA Class I	Class I >10% vs Class I <10%	0.99 (0.62, 1.59)
Hayes	2015		Pre-Transplant FVC	% precicted	1.00 (1.00, 1.00)
Demir	2015		CMV Status: Recipient	Negative vs Positive	1.00 (0.76, 1.32)
Hayes	2015		Pre-Transplant FEV1	% predicted	1.00 (1.00, 1.00)
George	2012		Pre-Transplant mPAP	per mmHg	1.00 (0.99, 1.01)
Shah	2008		PRA	per unit increase	1.00 (1.00, 1.01)
Chauhan	2016		Diastolic Blood Pressure	per mmHg	1.02 (1.01, 1.03)
Osho	2015		eGFR	per 1 unit decrease	1.02 (0.96, 1.08)
Hayes	2015		Pre-Transplant Systolic PAP	<35 mmHg vs >35 mmHg	1.03 (0.88, 1.21)
McKellar	2016		Bilirubin	mg/dL	1.05 (1.03, 1.07)
Osho	2015		eGFR Decrease	Decrease >25% from Baseline vs No Decrease	1.06 (0.97, 1.16)
Hayanga	2015		Bilirubin	Elevated vs Normal	1.08 (1.03, 1.13)
Hayes	2015	Age Threshold = 70	Creatinine	mg/dL	1.09 (1.04, 1.14)
Hayes	2015	Age Threshold = 60	Creatinine	mg/dL	1.09 (1.05, 1.14)
Awori	2016		eGFR	<60 vs >90	1.13 (1.03, 1.23)
Schaffer	2015		Pre-Transplant mPAP	>30 mmHg vs <30 mmHg	1.14 (1.00, 1.30)
Baldwin	2012		Albumin	3.7 - 3.9 vs 4.0 - 5.0	1.19 (1.01, 1.40)
Osho	2015		eGFR	>50 vs <50	1.27 (1.09, 1.48)
Baldwin	2012		Albumin	3.0 - 3.6 vs 4.0 - 5.0	1.33 (1.13, 1.57)
Shah	2008		PRA	<25 % vs >25%	1.35 (1.08, 1.69)
Osho	2015		eGFR Decrease	Decrease >50% from Baseline vs No Decrease	1.44 (1.28, 1.62)
Plantier	2010		Pre-Transplant Cardiac Index	L/min/m2	1.47 (1.13, 1.91)
Baldwin	2012		Albumin	0.5 g/dL decrease	1.48 (1.21, 1.81)
Baldwin	2012		Albumin	5.1 - 5.9 vs 4.0 - 5.0	1.50 (0.90, 2.49)
Baldwin	2012		Albumin	0.5 - 2.9 vs 4.0 - 5.0	1.55 (1.17, 2.05)
Awori	2016		eGFR	60-90 vs >90	1.61 (1.42, 1.83)
Osho	2015		eGFR	>40.2 vs < 40.2	1.72 (1.37, 2.15)
Hayes	2016		PRA Class II	Class II >10% vs Class II <10%	1.98 (1.17, 3.37)

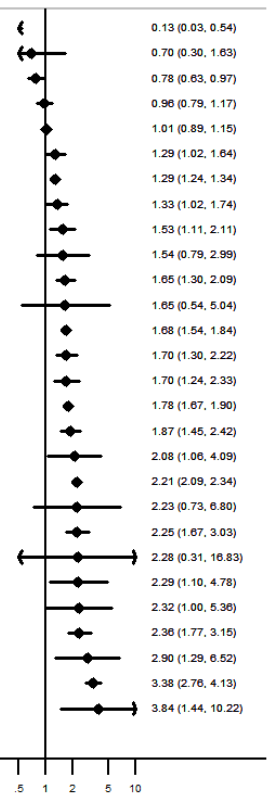


APPENDIX AP – List of unique predictors identified by one study related to recipient psychology/sociology



APPENDIX AQ – List of unique predictors identified by one study related to recipient psychology/sociology

First Author	Year	Stratified Group	Predictor	Category	Hazard Ratio (95% CI)
Atkins	2010		Swallowing Evaluation	Clear vs Failed	0.13 (0.03, 0.54)
Mihalek	2013		Interstitial Pneumontitis	Yes vs No Post-Tx	0.70 (0.30, 1.63)
O'Keefe	2014		mPAP Increase	>5 mmHg increase Post-Tx vs Unchanged Post-Tx	0.78 (0.63, 0.97)
O'Keefe	2014		mPAP Decrease	>5 mmHg decrease Post-Tx vs Unchanged Post-Tx	0.96 (0.79, 1.17)
Shigemura	2014		Mechanical Ventilation Time	>5 days Post-Tx vs <5 days Post-Tx	1.01 (0.89, 1.15)
Chan	2016		Cardiac Events	Yes vs No Post-Tx	1.29 (1.02, 1.64)
Shigemura	2014		Delayed Chest Closing	Chest Open >7 days Post-Tx vs <7 days Post-Tx	1.29 (1.24, 1.34)
Chan	2016		Vascular Events	Yes vs No Post-Tx	1.33 (1.02, 1.74)
Kopp	2015		Treatment for Rejection	Yes within one year Post-Tx vs No Post-Tx	1.53 (1.11, 2.11)
Jacques	2012		Acute Renal Failure	Yes vs No Post-Tx	1.54 (0.79, 2.99)
Hayes	2016	PRA Class I >10%	Bronchiolitis Obliterans Syndrome	Yes vs No Post-Tx	1.65 (1.30, 2.09)
Arnaoutakis	2011		Post-Operative Tracheostomy	Yes vs No Post-Tx	1.65 (0.54, 5.04)
Shigemura	2014		Delayed Chest Covering (Covered Sterile Bandanges)	Yes vs No Post-Tx	1.68 (1.54, 1.84)
Chan	2016		Renal Events	Yes vs No Post-Tx	1.70 (1.30, 2.22)
Wehbe	2012		AKI: AKIN Class I	AKIN Level 1 Post-Tx vs No AKI Post-Tx	1.70 (1.24, 2.33)
Shigemura	2014		Renal Insufficiency Post Transplant	Yes vs No Post-Tx	1.78 (1.67, 1.90)
Hayes	2016	PRA Class II >10%	Bronchiolitis Obliterans Syndrome	Yes vs No Post-Tx	1.87 (1.45, 2.42)
Atkins	2010		Mechanical Ventilation	Yes vs No Post-Tx	2.08 (1.06, 4.09)
Shigemura	2014		Delayed Chest Covering (Bandages and Retractor)	Yes vs No Post-Tx	2.21 (2.09, 2.34)
Minambres	2009		Suture Dehiscence	Yes vs No Post-Tx	2.23 (0.73, 6.80)
Shino	2013		Diffuse Alveolar Damage	Yes vs No Post-Tx	2.25 (1.67, 3.03)
Newton	2017		Acute Cellular Rejection Score	per one score unit increase	2.28 (0.31, 16.83)
Huppmann	2012		Cumulative Acute Rejections	>2 within first year vs <2	2.29 (1.10, 4.78)
Huppmann	2012		Lymphocytic Bronchiolitis Incidence	> 2 cases in first year Post-Tx vs < 2	2.32 (1.00, 5.36)
DerHovanessian	2016		Forced Vital Capacity Decrease	Yes vs No Post-Tx	2.36 (1.77, 3.15)
Wehbe	2012		AKI: AKIN Class II	AKIN Level 2-3 Post-Tx vs No AKI Post-Tx	2.90 (1.29, 6.52)
Awori	2016		Airway Complications	Yes vs No Post-Tx	3.38 (2.76, 4.13)
Henri	2012		Bronchial Dehiscence	Yes vs No Post-Tx	3.84 (1.44, 10.22)



Chapter 5: Risk Factors for One-Year Graft Loss Post-Kidney Transplantation: A systematic review and meta-analyses

STATUS: Published – Clinical Journal of American Society of Nephrology

https://doi.org/10.2215/CJN.05560519

Farid Foroutan[1,3], Erik Friesen[2], Kathryn Elizabeth Clark[1], Shahrzad Motaghipisheh[3], Roman Zyla[2], Yung Lee[3], Rakhshan Kamran[3], Emir Ali[3], Mitch De Snoo, Ani Orchanian-Cheff[5], Christine Ribic[4], Darin Treleaven[4], Gordon Guyatt[3], Maureen Meade[3]

1- Ted Rogers Centre for Heart Research, Multi-Organ Transplant Program, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

2- Department of Medicine, University of Toronto, Toronto, Ontario, Canada

3- Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

4 – Department of Medicine, McMaster University, Hamilton, Ontario, Canada

5 – Library and Information Services, University Health Network, Toronto, Ontario, Canada

ABSTRACT

Background

With expansion of the pool of kidney grafts, through the use of higher risk donors, and increased attention to donor management strategies, the 1-year graft survival rate is subject to change. It is, therefore, useful to elucidate 1-year graft survival rates by dissecting the characteristics of the low-risk and high-risk kidney transplant cases. The objective of our study was to evaluate factors purported to influence the risk of 1-year graft loss in kidney transplant recipients.

Methods

We searched bibliographic databases from 2000 to 2017 and included observational studies that measured the association between donor, recipient, the transplant operation, or early post-operative complications, and 1-year death-censored graft loss.

Results

We identified 35 eligible primary studies, with 20 risk factors amenable to meta-analysis. Six factors were associated with graft loss, with moderate to high degree of certainty: donor age (HR 1.11 per 10-year increase, 95% CI 1.04 to 1.18), extended criteria donors (HR 1.35, 95% CI 1.28 to 1.42), deceased donors (HR 1.54, 95% CI 1.32 to 1.82), number of human leukocyte antigen (HLA) mismatches (HR 1.08 per 1 mismatch increase, 95% CI 1.07 to 1.09), recipient age (HR 1.17 per 10-year increase, 95% CI 1.09 to 1.25), and delayed graft function (HR 1.89, 95% CI 1.46 to 2.47) as risk factors for 1-year graft loss. Pooled analyses also excluded, with a high degree of certainty, any associations of cold ischemia time, recipient race, pre-transplant BMI, diabetes, and hypertension with 1-year graft loss.

Conclusion

Recipient age, donor age, standard vs extended criteria donor, living vs deceased donor, HLA mismatch, and delayed graft function all predicted 1-year graft survival. The effect of each risk factor is small.

INTRODUCTION

In patients with end stage kidney disease receiving kidney replacement therapy transplant enormously improves quality of life and survival,[1] However, due to the high demand and limited supply of available kidneys, many patients will undergo dialysis for up to 11 years or more prior to kidney transplant.[2]

After transplantation, maximizing graft longevity becomes a focus of care. Graft loss results in return to dialysis, re-transplantation, or death. Kidney transplant recipients have the highest rate of graft survival among all organs transplanted: 92% 1-year graft survival for kidneys transplanted from deceased donors.[3] With expansion of the pool of kidney grafts, through the use of higher risk donors, and increased attention to donor management strategies, 1-year graft survival may change. It is therefore useful to identify low-risk and high-risk kidney transplant cases.

Prognostic studies can guide clinicians and patients in better understanding factors associated with a higher risk of graft loss in the first-year post-transplantation. Although formal risk prediction models can inform prognosis, existing models in kidney transplant perform poorly: the discriminatory performance of existing models ranges from 0.54 to 0.72, either below or marginally above the minimal threshold (0.6) for acceptable performance[4 5]. The limited performance of current models may result from including risk factors useful in one cohort but not in others due to varied management protocols across centers and over time, varied or suboptimal adjustment for covariates, or risk of bias in the primary studies. A systematic review and meta-analysis of studies assessing these factors improves the precision of their associations and allow for exploration of potential sources of discrepancy between studies. Because a systematic review and meta-analysis could guide the development of a prediction model with

useful discrimination and calibration[6], we undertook a review to assess the predictive power of key risk factors for kidney graft survival at one year post transplant.

METHODS

Data sources and searches

With the help of an information specialist we searched bibliographic databases in February of 2017 (supplemental material). Specifically, we searched MEDLINE, EMBASE, Cochrane central register for controlled trials, and Cochrane database for systematic reviews for citations between the years 2000 to 2017.

Study selection and data extraction

The supplemental material provides details of the selection process and data extraction. Briefly, we selected observational studies of adult (≥18 years) kidney recipients receiving their first transplant, including studies evaluating the association between any risk factors and 1-year graft-loss using multivariable analysis. We did not restrict by language or publication status. We included identified abstracts that met our inclusion criteria and provided enough information to contribute to our study. We also relied on the expertise of our clinical experts to inform us of any unpublished data not captured by our search strategy. From the final set of eligible studies, data abstractors recorded data from each study directly into a structured and pre-tested excel database.

Risk of bias of individual studies

We assessed the risk of bias of individual studies using the Quality in Prognostic Studies (QUIPS) instrument[7]. When we judged 5 or more of the 6 QUIPS domains to be at low risk of bias, we classified the overall risk of bias as low; otherwise we considered at high risk of bias.

Data synthesis and statistical analysis

We conducted meta-analysis for any risk factor evaluated in two or more studies. When a risk factor was addressed by only one study, we present the reported point estimate and 95% CI. The included studies reported point estimates and 95% confidence intervals (CIs) as hazard ratios (HR), odds ratios (OR) or relative risks (RR). Due to the low risk of graft loss within the first year following transplantation, we included OR and RR in the same meta-analysis without conversion [8 9]. To combine studies that reported HR with those reporting OR or RR, we conducted sub-group comparisons. When we observed a clinically or statistically significant difference between binary (e.g., OR or RR) and time-to-event measures (e.g., HR), we converted the OR or RR to HR using baseline risk estimates from the individual studies. When studies did not provide baseline risks, we utilized the average risk, prevalence of the risk factor, and the relative effect to estimate the baseline risk[10]. The supplemental material includes further details of the data synthesis.

We addressed statistical heterogeneity through visual inspection of forest plots, looking for the consistency of point estimates and the extent of overlap in confidence intervals. Heterogeneity was not assessed with the I^2 statistic, which is not useful in observational studies with a very large sample size [11].

This review addressed two possible subgroup analyses: risk of bias and outcome definition. The supplemental material presents our hypotheses for these two subgroup analyses.

When the sub-group analysis for risk of bias and outcome definition showed a significant difference across groups, we focused the analysis on studies at low risk of bias and/or those assessing death censored graft failure and applied the GRADE assessment only to these studies.

We applied a two-sided P value of 0.05 or less to denote statistical significance. STATA's *metan* function provided the platform for conducting all statistical analyses[12].

Certainty in the body of evidence

To assess the certainty of evidence across all studies related to a given risk factor, we used GRADE approach that rates the certainty of evidence as high, moderate, low, or very low considering issues of risk of bias, imprecision, inconsistency, indirectness and publication bias[11]. We assessed publication bias using visual inspection of funnel plots.

RESULTS

Study selection and characteristics

The literature search identified 19,679 unique citations of which 2,220 citations required full text review; 35 studies ultimately proved eligible[13-47]. Appendix A provides a summary of study characteristics. The individual studies included patients from Canada, Denmark, Germany, Ireland, Italy, Japan, Norway, Portugal, Spain, South Korea, Taiwan, United Kingdom, and United States.

Risk of bias of individual studies

Of the 35 eligible studies, reviewers judged 18 to have high risk of bias (appendix B).[13 18 19 21-23 25 27 28 35-42 45] most commonly because of limitations in statistical analysis (over fitting of the regression models, building a multivariable model based on level of significance in univariable analysis, and inclusion of collinear variables) and reporting (such as only reporting on the significant risk factors). Amongst the included studies, the authors included an average of 11 variables (standard deviation of 6, minimum of 3 and maximum of 23). Across the many risk factors included in this review, only the subgroup analyses for risk of bias in recipient diabetes and delayed graft function showed statistically significant different effect estimates in studies at high versus low risk of bias. For these, we only utilized estimates from low risk of bias studies.

Meta-analyses of Donor Factors

The review assessed 6 donor characteristics; 5 were independently associated with one-year graft loss in the original studies and also proved predictive in the meta-analysis (Table 1): donor type (HR 1.54 for deceased donors, 95% CI 1.32 to 1.82; high certainty), donor quality (HR 1.35 for extended-criteria donors, 95% CI 1.28 to 1.42; moderate certainty due to risk of bias), donor age (HR 1.11 per 10-year increase, 95% CI 1.04 to 1.18; high certainty), donor sex (HR 1.10 for female sex, 95% CI 1.07 to 1.21; moderate certainty, due to serious inconsistency), and donor body mass index (BMI) (HR 0.90 per 10 kg/m^2 increase, 95% CI 0.82 to 0.91; moderate certainty due to serious risk of bias). We observed that all studies defined extended criteria donors as >60 years of age or age 50 to 59 years with two of three associated risk factors—history of cerebrovascular accident, hypertension, or serum creatinine greater than 1.5 mg/dL and delayed graft function as the need for dialysis within the first week post-transplant. We did not observe a statistically significant association between donor serum creatinine level and the risk of 1-year graft loss (table 1). We did not detect publication bias for any of the donor factors.

Meta-analyses of Transplant Process Factors

We assessed two risk factor variables characteristic of the transplant process (Table 1, Figure 2). The number of HLA mismatches was the only risk factor, for which we observed an association beyond chance, with 1-year death-censored graft loss (HR 1.08 per 1-mismatch increase, 95% CI 1.07 to 1.09; high certainty). We observed no significant association between cold ischemia time (HR 1.001 per 1-hour increase, 95% CI 0.998 to 1.004) and graft loss, despite 5 studies evaluating this variable, adjusted for recipient age, donor age[S6, S18, S23, S28, S30], donor sex[S28], donor cause of death[S18], donor type[S6, S23, S30], HLA mismatch[S6, S18, S28, S30], recipient sex[S6, S23, S28], recipient BMI[S18], recipient diabetes[S6, S23, S42], pre-transplant time on dialysis[S6, S18, S23, S30], history of cardiovascular

comorbidities[S6, S23], delayed graft function[S6, S18], and early acute rejection[S6, S28, S30]. We did not detect publication bias for any of the transplant process factors.

Meta-analyses of Recipient Factors

We identified nine transplant recipient variables that had been investigated in two or more of the primary studies in this review (Table 1, Figure 2). Four of the nine were significantly associated with 1-year death-censored graft loss: recipient age (HR 1.17 per 10-year increase, 95% CI 1.09 to 1.25; high certainty), pre-transplant smoking (HR 1.59, 95% CI 1.34 to 1.90; moderate certainty due to serious imprecision), pre-transplant recipient coronary artery disease (HR 1.15, 95% CI 1.03 to 1.27; moderate certainty due to serious indirectness), and number of pre-transplant years on dialysis (HR 1.03 per 1-year increase, 95% CI 1.02 to 1.03; moderate certainty due to serious risk of bias). We did not observe a statistically significant association for recipient sex, race, BMI, hypertension, or diabetes with one-year graft loss (Table 1, Figure 2). We did not detect publication bias for any of the recipient factors.

Meta-analyses of Post-transplant Complications

The literature included within this review commonly identified delayed graft function and acute rejection as early post-transplant complications associated with death-censored graft loss (Table 1, Figure 2). For both, we observed a statistically significant association with 1-year graft loss: delayed graft function (HR 1.89, 95% CI 1.46 to 2.47; moderate certainty due to serious inconsistency); acute rejection (HR 3.16, 95% CI 1.86 to 5.38; moderate certainty due to serious inconsistency). We did not detect publication bias for any of the post-transplant complication factors.

Risk Factors Addressed in a Single Study

This review identified an additional 72 candidate risk factors, each evaluated in only one study. We summarized the full list of risk factors in the appendix.

DISCUSSION

Principal findings

This review identified 5 risk factors, for which there is moderate to high certainty in the magnitude of association with one-year graft loss: donor age, extended criteria donors, deceased donors, increasing number of HLA mismatches, and recipient age. We identified an additional five variables for which, with moderate certainty, there is an association with 1-year graft loss: donor sex, donor BMI, recipient's number of years on dialysis, history of smoking, and coronary artery disease. With high certainty, the findings of this study exclude any association of the following variables with one-year graft loss: increasing cold ischemia time, recipient age, recipient BMI, recipient diabetes, and recipient hypertension.

Strengths and Limitations

This study is the first large-scale systematic review of studies that have conducted adjusted analyses addressing risk factors for 1-year graft loss after kidney transplantation. By only reviewing adjusted evidence, users of our estimates can multiple the HR of multiple risk factors to obtain their combined effect on the risk of 1-year graft loss. Using rigorous meta-analytic methods, the review provides precise measures, compared to any individual study, for the association of each risk factor and graft loss, informed by observational cohort studies. The use of GRADE methodology enabled us to not only report on the direction and magnitude of the association for each risk factor, but also to transparently report on the certainty of the evidence.

One limitation of this review is that we included studies identifying risk factors using Cox regression analysis for graft loss at all time points in follow-up. By doing so, we assumed that the authors of the primary studies had tested and ensured the proportional hazards assumption necessary for validity of any reported hazard ratio. The authors of the individual studies seldom reported on assessing the necessary assumptions of their regression models. As a result, our statistical analysis for the risk of bias assessment could not be fully informed by meeting the regression model assumptions.

We included UNOS registry studies to represent all studies published from individual centers in the USA. By doing so, the quality of this review is dependent upon the quality of the UNOS registry data. Authors of single- or multi-centered observational studies may have more direct control over their data collection and entry compared to large registries[48], and thus more likely to ensure data quality prior to analysis of risk factors. We utilized evidence from the UNOS registry as this is the source that is highly referred to by the transplant community.

Studies varied considerably in the covariates included in their predictive models (studies included 11$\pm$6 covariates in their regression models). Thus, results are vulnerable to the possibility that the impact of a particular risk factor might differ depending on which variables were included in a particular model.

In the context of identifying factors that increase the risk of graft loss in the year following transplant, the studies in this review have a fundamental limitation: potential candidates for transplant may be rejected because of patient factors that were not included. The reasons for not recommending transplant in such individuals may be the most powerful determinants of outcome. These may include, but not be limited to: active infections, combination of older age

with constellation of other comorbidities such as obesity, cardiovascular disease, malignancies, and irreversible obstructive or restrictive pulmonary disease[49].

Some may be surprised that cold ischemia time was not associated with 1-year graft loss. The simplest explanation for this finding is that, indeed, there is no association. Another plausible explanation is that present-day use of storage techniques such as machine perfusion and preservation solutions minimize cold-ischemia damage to kidney[50 51]. Additionally, studies treated cold ischemia time as a continuous variable and assumed a linear relationship between ischemia time and survival. It is possible that the relationship is non-linear. For instance, up to a certain duration, there may be no relation between ischemia time and outcome, but beyond that duration graft longevity diminshes[S1, S3, S8,13 15 20 52]. Amongst the risk factors addressed by individual studies, we identified two that treated ischemia time as a binary variable. One studied used the threshold of 20 hours and observed a HR of 1.92 (95% CI 1.26 – 2.91)[S1]. The other used a threshold of 24 hours and observed a HR of 1.27 (1.09 – 1.48)[S3]. Both of these studies suggest that ischemia time is associated with 1-year graft loss only after a long passage of time. Therefore, assuming a linear association might have put primary studies at high risk of missing such a non-linear relationship.

Relation to other work

Kaboré et al. conducted a systematic review and meta-analysis to identify all risk prediction models for graft loss post kidney transplantation[4]. Of the 34 identified models, only 7[53-59] specifically predict graft loss at 1-year post-transplantation. The median discrimination value, as measured by area under the curve (AUC) statistics, is 0.63 (range from 0.54 to 0.72). One potential reason for poor discrimination is that these models use risk factors that we identified

not to be associated with 1-year graft loss beyond chance. For example, Tang et al. included recipient sex, race, height, weight, diabetes, history of hypertension, and cold ischemic time within their model to predict graft-loss at 1-year post transplantation[58]. Our review excluded an association beyond chance for each of these risk factors and 1-year graft loss. This is one plausible explanation for Tang et al. observing an AUC of 0.54.

Contrary to their inclusion within the aforementioned risk prediction models, our review excluded an association beyond chance for factors including recipient sex (present in 4 of the 7 aforementioned risk prediction models), race (5 of 7 models), BMI (4 of 7 models), diabetes (4 of 7 models), and hypertension (4 of 7 models), donor creatinine (2 of 7 models), and cold ischemia time (5 of 7 models). Previous studies reported that female recipients have better long-term prognosis compared to men. Such better prognosis has been hypothesized to be due to hormonal protection[60]. Such biologic explanations may require longer duration of follow-up (beyond 1-year) to express their effect. This review's short follow-up time of 1-year may be the reason for lack of association beyond chance for recipient sex and graft loss. This review's finding of a lack of association between race and graft loss can be explained by the diminishing racial disparity in kidney transplantation. Recent analysis of the Scientific Registry of Transplant Recipients (SRTR) registry from the US suggests significant improvement in graft survival from 1990 to 2012, with the success rate improving in black recipients to a greater extent than improvement observed in white recipients[61].

Of all the factors not associated with graft-loss, we were most surprised to find no evidence of an association between donor creatinine and 1-year graft loss. Both of the studies in this review that evaluated the prognostic importance of donor creatinine utilized non-death censored graft

loss as their outcome. The inclusion of patient mortality may explain the lack of association between donor creatinine and mortality (it is possible that worse functioning kidneys would not be associated with patient mortality due to the availability of kidney replacement therapies in the event of graft failure). Another explanation may be that donors with high creatinine were not selected for transplantation, thus eliminating any association beyond chance. The association between donor creatinine and 1-year graft loss is partially captured by the significant association between extended criteria donor as a risk factor for 1-year graft loss. Additionally, amongst the risk factors addressed in only one study, we identified donor eGFR >60 ml/min/kg to be associated with a decrease in risk of graft loss at 1-year post kidney transplantation (supplemental figure 1).

Implication for guidelines

From this review, however, it is evident that numerous recipient and donor characteristics increase the risk of graft loss post kidney transplantation. All such factors, although may be associated with graft loss beyond chance, may not be clinically important to diminish the magnitude of benefit attained from transplantation. This necessitates the need for risk prediction models to guide clinicians in selection of candidates whose risk for graft loss (disadvantaging the societal need for organ donors), may be higher than their risk of mortality on dialysis. Risk associations generated from this review may inspire or provide the foundational information necessary for development of a risk prediction model.

CONLUSION

Our systematic review and meta-analysis identified 10 risk factors for which we have moderate or high certainty in their strength and magnitude of association. These factors include recipient

age, donor age, extended criteria donors, deceased donors, and increasing number of HLA mismatches. With high certainty, we were able to establish that increasing cold ischemia time, recipient's BMI, recipient diabetes, and recipient hypertension do not have large associations with one-year graft survival. The optimal utilization of the factors we have identified as risk factors, in development of future risk prediction models, may improve discrimination and calibration. Such models in turn may guide the judgment clinicians need to make on the highest risk recipient and donor.

REFERENCES

1. Tonelli M, Wiebe N, Knoll G, et al. Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. *Am J Transplant* 2011;11(10):2093-109. doi: 10.1111/j.1600-6143.2011.03686.x [published Online First: 2011/09/03]
2. Matas AJ, Smith JM, Skeans MA, et al. OPTN/SRTR 2013 Annual Data Report: kidney. *Am J Transplant* 2015;15 Suppl 2:1-34. doi: 10.1111/ajt.13195 [published Online First: 2015/01/30]
3. Hart A, Smith JM, Skeans MA, et al. OPTN/SRTR 2016 Annual Data Report: Kidney. *Am J Transplant* 2018;18 Suppl 1:18-113. doi: 10.1111/ajt.14557 [published Online First: 2018/01/03]
4. Kabore R, Haller MC, Harambat J, et al. Risk prediction models for graft failure in kidney transplantation: a systematic review. *Nephrol Dial Transplant* 2017;32(suppl_2):ii68-ii76. doi: 10.1093/ndt/gfw405 [published Online First: 2017/02/17]
5. Alba AC, Agoritsas T, Walsh M, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA* 2017;318(14):1377-84. doi: 10.1001/jama.2017.12126 [published Online First: 2017/10/20]
6. Alba AC, Walter SD, Guyatt GH, et al. Predicting Survival in Heart Failure Patients With Implantable Cardioverter-Defibrillators: The Heart Failure Meta-score. *J Card Fail* 2017 doi: 10.1016/j.cardfail.2017.11.002 [published Online First: 2017/11/28]
7. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6. doi: 10.7326/0003-4819-158-4-201302190-00009 [published Online First: 2013/02/20]
8. Grant RL. Converting an odds ratio to a range of plausible relative risks for better communication of research findings. *BMJ* 2014;348:f7450. doi: 10.1136/bmj.f7450 [published Online First: 2014/01/28]
9. Tierney JF, Stewart LA, Ghersi D, et al. Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials* 2007;8:16. doi: 10.1186/1745-6215-8-16 [published Online First: 2007/06/09]
10. Kooter AJ, Kostense PJ, Groenewold J, et al. Integrating information from novel risk factors with calculated risks: the critical impact of risk factor prevalence. *Circulation* 2011;124(6):741-5. doi: 10.1161/CIRCULATIONAHA.111.035725 [published Online First: 2011/08/10]
11. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;350:h870. doi: 10.1136/bmj.h870 [published Online First: 2015/03/18]
12. Harris R, Bradburn M, Deeks J, et al. metan: fixed- and random-effects meta-analysis. *Stata Journal* 2008;8(1):3-28.
13. Asderakis A, Dyer P, Augustine T, et al. Effect of cold ischemic time and HLA matching in kidneys coming from "young" and "old" donors: do not leave for tomorrow what you can do tonight. *Transplantation* 2001;72(4):674-8. [published Online First: 2001/09/07]
14. Lin HH, Huang CC, Huang JY, et al. Impact of HCV infection on first cadaveric renal transplantation, a single center experience. *Clin Transplant* 2004;18(3):261-6. doi: 10.1111/j.1399-0012.2004.00153.x [published Online First: 2004/05/15]

15. Andresdottir MB, Haasnoot GW, Doxiadis, II, et al. Exclusive characteristics of graft survival and risk factors in recipients with immunoglobulin A nephropathy: a retrospective analysis of registry data. *Transplantation* 2005;80(8):1012-8. [published Online First: 2005/11/10]
16. Cardinal H, Hebert MJ, Rahme E, et al. Modifiable factors predicting patient survival in elderly kidney transplant recipients. *Kidney Int* 2005;68(1):345-51. doi: 10.1111/j.1523-1755.2005.00410.x [published Online First: 2005/06/16]
17. Courtney AE, McNamee PT, Middleton D, et al. Association of functional heme oxygenase-1 gene promoter polymorphism with renal transplantation outcomes. *Am J Transplant* 2007;7(4):908-13. doi: 10.1111/j.1600-6143.2006.01726.x [published Online First: 2007/03/30]
18. Kruger B, Boger CA, Schroppel B, et al. Impact of NOD2/CARD15 haplotypes on the outcome after kidney transplantation. *Transpl Int* 2007;20(7):600-7. doi: 10.1111/j.1432-2277.2007.00486.x [published Online First: 2007/05/15]
19. Diaz JM, Gich I, Bonfill X, et al. Prevalence evolution and impact of cardiovascular risk factors on allograft and renal transplant patient survival. *Transplant Proc* 2009;41(6):2151-5. doi: 10.1016/j.transproceed.2009.06.134 [published Online First: 2009/09/01]
20. Ferrer F, Mota A, Alves R, et al. Renal transplantation with expanded criteria donors: the experience of one Portuguese center. *Transplant Proc* 2009;41(3):791-3. doi: 10.1016/j.transproceed.2009.02.044 [published Online First: 2009/04/21]
21. Heldal K, Hartmann A, Leivestad T, et al. Clinical outcomes in elderly kidney transplant recipients are related to acute rejection episodes rather than pretransplant comorbidity. *Transplantation* 2009;87(7):1045-51. doi: 10.1097/TP.0b013e31819cdddd [published Online First: 2009/04/09]
22. Kayler LK, Garzon P, Magliocca J, et al. Outcomes and utilization of kidneys from deceased donors with acute kidney injury. *Am J Transplant* 2009;9(2):367-73. doi: 10.1111/j.1600-6143.2008.02505.x [published Online First: 2009/01/31]
23. Lynch RJ, Ranney DN, Shijie C, et al. Obesity, surgical site infection, and outcome following renal transplantation. *Ann Surg* 2009;250(6):1014-20. doi: 10.1097/SLA.0b013e3181b4ee9a [published Online First: 2009/09/26]
24. Fuggle SV, Allen JE, Johnson RJ, et al. Factors affecting graft and patient survival after live donor kidney transplantation in the UK. *Transplantation* 2010;89(6):694-701. doi: 10.1097/TP.0b013e3181c7dc99 [published Online First: 2010/03/24]
25. Moore J, McKnight AJ, Simmonds MJ, et al. Association of caveolin-1 gene polymorphism with kidney transplant fibrosis and allograft failure. *JAMA* 2010;303(13):1282-7. doi: 10.1001/jama.2010.356 [published Online First: 2010/04/08]
26. Papalia T, Greco R, Lofaro D, et al. Impact of body mass index on graft loss in normal and overweight patients: retrospective analysis of 206 renal transplants. *Clin Transplant* 2010;24(6):E241-6. doi: 10.1111/j.1399-0012.2010.01258.x [published Online First: 2010/05/21]
27. Heldal K, Hartmann A, Leivestad T, et al. Risk variables associated with the outcome of kidney recipients >70 years of age in the new millennium. *Nephrol Dial Transplant* 2011;26(8):2706-11. doi: 10.1093/ndt/gfq786 [published Online First: 2011/02/03]

28. Molnar MZ, Kovesdy CP, Bunnapradist S, et al. Associations of pretransplant serum albumin with post-transplant outcomes in kidney transplant recipients. *Am J Transplant* 2011;11(5):1006-15. doi: 10.1111/j.1600-6143.2011.03480.x [published Online First: 2011/04/01]
29. Carrier M, Lize JF, Quebec-Transplant P. Impact of expanded criteria donors on outcomes of recipients after kidney transplantation. *Transplant Proc* 2012;44(7):2227-30. doi: 10.1016/j.transproceed.2012.07.120 [published Online First: 2012/09/15]
30. Grosso G, Corona D, Mistretta A, et al. The role of obesity in kidney transplantation outcome. *Transplant Proc* 2012;44(7):1864-8. doi: 10.1016/j.transproceed.2012.06.043 [published Online First: 2012/09/15]
31. Molnar MZ, Streja E, Kovesdy CP, et al. Age and the associations of living donor and expanded criteria donor kidneys with kidney transplant outcomes. *Am J Kidney Dis* 2012;59(6):841-8. doi: 10.1053/j.ajkd.2011.12.014 [published Online First: 2012/02/07]
32. Nanmoku K, Matsuda Y, Yamamoto T, et al. Clinical characteristics and outcomes of renal transplantation in elderly recipients. *Transplant Proc* 2012;44(1):281-3. doi: 10.1016/j.transproceed.2012.01.006 [published Online First: 2012/02/09]
33. Andreoni KA, Forbes R, Andreoni RM, et al. Age-related kidney transplant outcomes: health disparities amplified in adolescence. *JAMA Intern Med* 2013;173(16):1524-32. doi: 10.1001/jamainternmed.2013.8495 [published Online First: 2013/07/31]
34. Bay JT, Schejbel L, Madsen HO, et al. Low C4 gene copy numbers are associated with superior graft survival in patients transplanted with a deceased donor kidney. *Kidney Int* 2013;84(3):562-9. doi: 10.1038/ki.2013.195 [published Online First: 2013/05/30]
35. Brar A, Jindal RM, Elster EA, et al. Effect of peripheral vascular disease on kidney allograft outcomes: a study of U.S. Renal data system. *Transplantation* 2013;95(6):810-5. doi: 10.1097/TP.0b013e31827eef36 [published Online First: 2013/01/29]
36. Faravardeh A, Eickhoff M, Jackson S, et al. Predictors of graft failure and death in elderly kidney transplant recipients. *Transplantation* 2013;96(12):1089-96. doi: 10.1097/TP.0b013e3182a688e5 [published Online First: 2013/09/24]
37. Nee R, Jindal RM, Little D, et al. Racial differences and income disparities are associated with poor outcomes in kidney transplant recipients with lupus nephritis. *Transplantation* 2013;95(12):1471-8. doi: 10.1097/TP.0b013e318292520e [published Online First: 2013/05/18]
38. Lee SH, Oh CK, Shin GT, et al. Age matching improves graft survival after living donor kidney transplantation. *Transplant Proc* 2014;46(2):449-53. doi: 10.1016/j.transproceed.2013.10.049 [published Online First: 2014/03/25]
39. Anderson JE, Steiner RW, Mekeel KL, et al. ECD kidney transplantation outcomes are improved when matching donors to recipients using a novel creatinine clearance match ratio (CCMR). *Clin Transplant* 2015;29(9):738-46. doi: 10.1111/ctr.12555 [published Online First: 2015/04/29]
40. Dinis P, Nunes P, Marconi L, et al. Small kidneys for large recipients: does size matter in renal transplantation? *Transplant Proc* 2015;47(4):920-5. doi: 10.1016/j.transproceed.2015.03.035 [published Online First: 2015/06/04]

41. Ilori TO, Adedinsewo DA, Odewole O, et al. Racial and Ethnic Disparities in Graft and Recipient Survival in Elderly Kidney Transplant Recipients. *J Am Geriatr Soc* 2015;63(12):2485-93. doi: 10.1111/jgs.13845 [published Online First: 2015/12/15]
42. Koo EH, Jang HR, Lee JE, et al. The impact of early and late acute rejection on graft survival in renal transplantation. *Kidney Res Clin Pract* 2015;34(3):160-4. doi: 10.1016/j.krcp.2015.06.003 [published Online First: 2015/10/21]
43. Adekoya AO, Halawa A. Kidneys From Deceased Elderly Donors: Factors Associated With Adverse Outcomes. *Exp Clin Transplant* 2016;14(1):32-7. [published Online First: 2016/02/11]
44. An JN, Ahn SV, Lee JP, et al. Pre-Transplant Cardiovascular Risk Factors Affect Kidney Allograft Survival: A Multi-Center Study in Korea. *PLoS One* 2016;11(8):e0160607. doi: 10.1371/journal.pone.0160607 [published Online First: 2016/08/09]
45. Huaman MA, Vilchez V, Mei X, et al. Donor positive blood culture is associated with delayed graft function in kidney transplant recipients: a propensity score analysis of the UNOS database. *Clin Transplant* 2016;30(4):415-20. doi: 10.1111/ctr.12703 [published Online First: 2016/02/04]
46. Redfield RR, Scalea JR, Zens TJ, et al. The mode of sensitization and its influence on allograft outcomes in highly sensitized kidney transplant recipients. *Nephrol Dial Transplant* 2016;31(10):1746-53. doi: 10.1093/ndt/gfw099 [published Online First: 2016/07/09]
47. Boffa C, van de Leemkolk F, Curnow E, et al. Transplantation of Kidneys From Donors With Acute Kidney Injury: Friend or Foe? *Am J Transplant* 2017;17(2):411-19. doi: 10.1111/ajt.13966 [published Online First: 2016/07/19]
48. Ferreira-Gonzalez I, Marsal JR, Mitjavila F, et al. Patient registries of acute coronary syndrome: assessing or biasing the clinical real world data? *Circ Cardiovasc Qual Outcomes* 2009;2(6):540-7. doi: 10.1161/CIRCOUTCOMES.108.844399 [published Online First: 2009/12/25]
49. Chadban SJ, Knoll GA, Ahn C, et al. KDIGO clinical practice guideline on the evaluation and the management of candidates for kidney transplantation: KDIGO; 2018 [Available from: https://kdigo.org/wp-content/uploads/2018/08/KDIGO-Txp-Candidate-GL-Public-Review-Draft-Oct-22.pdf.
50. Moers C, Smits JM, Maathuis MH, et al. Machine perfusion or cold storage in deceased-donor kidney transplantation. *N Engl J Med* 2009;360(1):7-19. doi: 10.1056/NEJMoa0802289 [published Online First: 2009/01/02]
51. Deng R, Gu G, Wang D, et al. Machine perfusion versus cold storage of kidneys derived from donation after cardiac death: a meta-analysis. *PLoS One* 2013;8(3):e56368. doi: 10.1371/journal.pone.0056368 [published Online First: 2013/03/29]
52. Postalcioglu M, Kaze AD, Byun BC, et al. Association of Cold Ischemia Time With Acute Renal Transplant Rejection. *Transplantation* 2018;102(7):1188-94. doi: 10.1097/TP.0000000000002106 [published Online First: 2018/01/19]
53. Bang K, Lee HK, Huh W, et al. Assessment of deceased donor kidneys using a donor scoring system. *Yonsei Med J* 2010;51(6):870-6. doi: 10.3349/ymj.2010.51.6.870 [published Online First: 2010/09/30]

54. Brown TS, Elster EA, Stevens K, et al. Bayesian modeling of pretransplant variables accurately predicts kidney graft survival. *Am J Nephrol* 2012;36(6):561-9. doi: 10.1159/000345552 [published Online First: 2012/12/12]
55. Krikov S, Khan A, Baird BC, et al. Predicting kidney transplant survival using tree-based modeling. *ASAIO J* 2007;53(5):592-600. doi: 10.1097/MAT.0b013e318145b9f7 [published Online First: 2007/09/22]
56. Lin RS, Horn SD, Hurdle JF, et al. Single and multiple time-point prediction models in kidney transplant outcomes. *J Biomed Inform* 2008;41(6):944-52. doi: 10.1016/j.jbi.2008.03.005 [published Online First: 2008/04/30]
57. Schold JD, Kaplan B, Baliga RS, et al. The broad spectrum of quality in deceased donor kidneys. *Am J Transplant* 2005;5(4 Pt 1):757-65. doi: 10.1111/j.1600-6143.2005.00770.x [published Online First: 2005/03/12]
58. Tang H, Hurdle JF, Poynton M, et al. Validating prediction models of kidney transplant outcome using single center data. *ASAIO J* 2011;57(3):206-12. doi: 10.1097/MAT.0b013e3182121bc5 [published Online First: 2011/03/11]
59. Kasiske BL, Israni AK, Snyder JJ, et al. A simple tool to predict outcomes after kidney transplant. *Am J Kidney Dis* 2010;56(5):947-60. doi: 10.1053/j.ajkd.2010.06.020 [published Online First: 2010/08/31]
60. Puoti F, Ricci A, Nanni-Costa A, et al. Organ transplantation and gender differences: a paradigmatic example of intertwining between biological and sociocultural determinants. *Biol Sex Differ* 2016;7:35. doi: 10.1186/s13293-016-0088-4 [published Online First: 2016/07/30]
61. Purnell TS, Luo X, Kucirka LM, et al. Reduced Racial Disparity in Kidney Transplant Outcomes in the United States from 1990 to 2012. *J Am Soc Nephrol* 2016;27(8):2511-8. doi: 10.1681/ASN.2015030293 [published Online First: 2016/02/06]

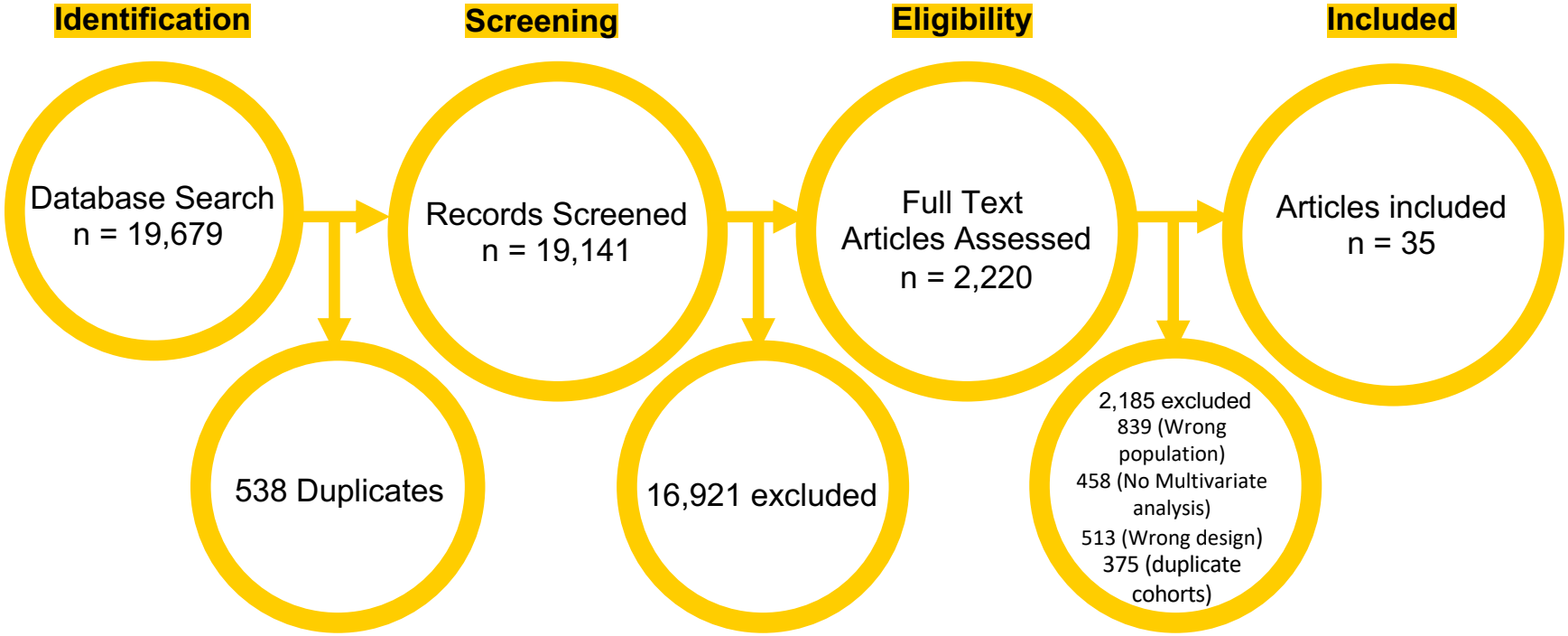
ACKNOWELDGEMENTS / FUNDING

This study received no funding.

DISCLOSURES

The authors have no financial or intellectual conflicts of interest to declare.

Figure 1 - PRISMA flow diagram for database search results and study selection



Wrong design: randomized controlled trials or studies, case-control studies, case-series, or editorials/commentaries.
Wrong population: Excluding patients with early graft survival (any study that excluded patients with graft survival less than 1-year) .
Recipients of multi-organ transplants or undergoing re-transplantation.
Duplicate cohorts: two studies with the same cohort reporting on the same predictor and exact same outcome.

Figure 2 – The association of all prognostic factors and 1-year graft loss identified and meta-analyzed

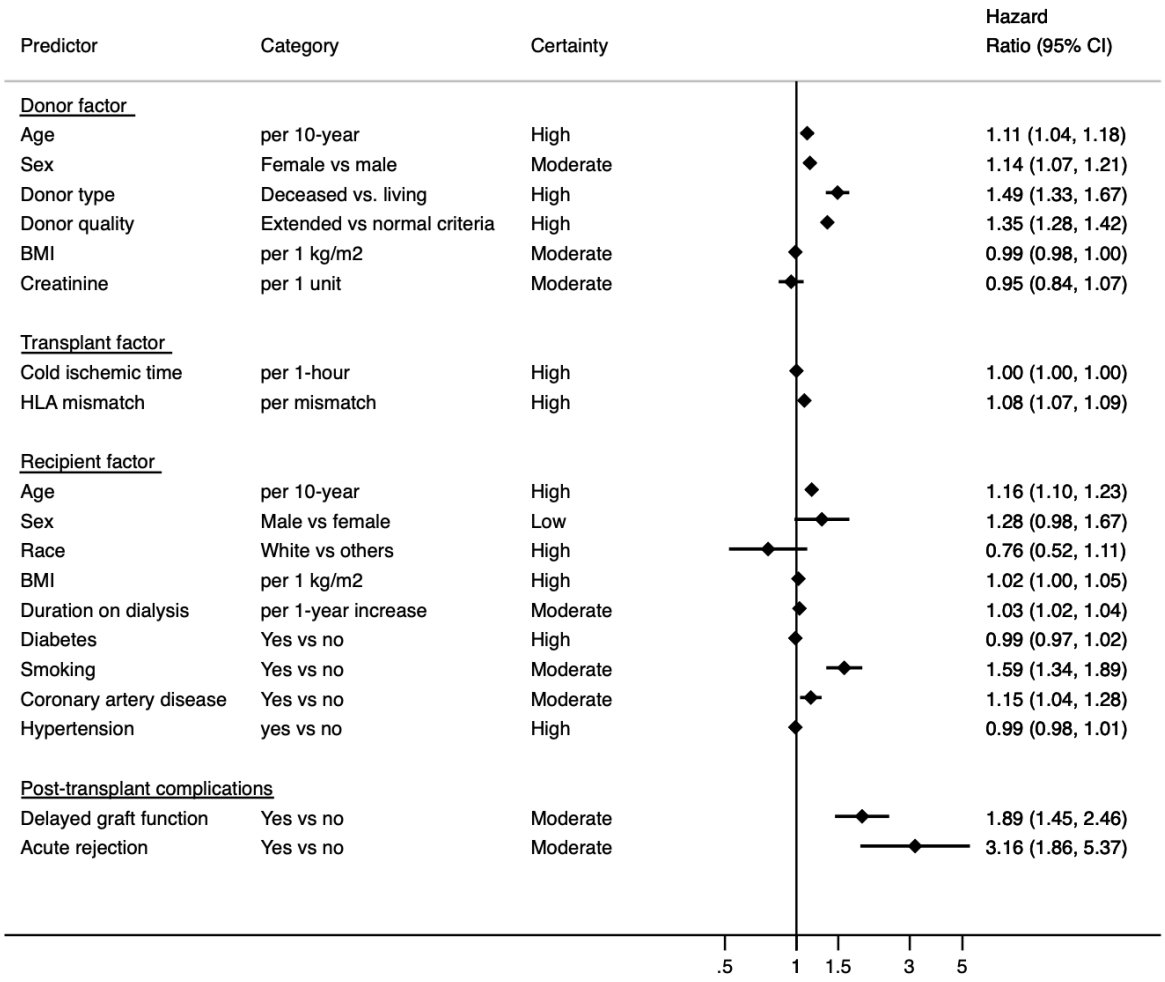


Table 2 - Summary of findings table for all risk factors commonly identified amongst included studies

Risk Factor	Study Results and measurements	Absolute effect estimates Baseline	Absolute effect estimates With predictor	Certainty in effect estimates (Quality of evidence)	Plain text summary
Donor Characteristics					
Age (10-year increase)	Hazard Ratio 1.11 (CI 95% 1.04 - 1.18) Based on data from 178,043 patients in 9 studies	74 per 1000	82 per 1000	High	Donor age increases graft failure
		Difference: 8 more per 1000 (CI 95% 3 more - 13 more)			
Sex (female vs male)	Hazard Ratio 1.10 (CI 95% 1.07 - 1.21) Based on data from 160,830 patients in 7 studies	70 per 1000	80 per 1000	Moderate Due to serious inconsistency	Female sex probably increases the risk for graft failure slightly
		Difference: 10 more per 1000 (CI 95% 5 more - 14 more)			
Type (deceased vs living)	Hazard Ratio 1.49 (CI 95% 1.33 - 1.67) Based on data from 149,433 patients in 4 studies	60 per 1000	90 per 1000	High	Deceased donors increase graft failure
		Difference: 30 more per 1000 (CI 95% 21 more - 38 more)			
Extended criteria donors (Yes vs no)	Hazard Ratio 1.35 (CI 95% 1.28 - 1.42) Based on data from 145,879 patients in 2 studies	72 per 1000	97 per 1000	Moderate Due to serious risk of bias	Extended criteria donor slightly increases graft failure
		Difference: 25 more per 1000 (CI 95% 20 more - 30 more)			
BMI (1 kg/m2 increase)	Hazard Ratio 0.99 (CI 95% 0.98 - 099) Based on data from 51,249 patients in 2 studies	74 per 1000	73 per 1000	Moderate Due to serious risk of bias	Higher Donor BMI probably decreases graft failure slightly
		Difference: 1 more per 1000 (CI 95% 1 fewer - 1 more)			
Creatinine (1 unit increase)	Hazard Ratio 0.95 (CI 95% 0.84 - 1.07) Based on data from 52,423 patients in 2 studies	74 per 1000	70 per 1000	Moderate Due to serious indirectness	Donor creatinine probably has little or no difference on graft failure
		Difference: 4 fewer per 1000 (CI 95% 11 fewer - 5 more)			
Transplant Characteristics					
Cold ischemic time (1 hour increase)	Hazard Ratio 1.00 (CI 95% 1.00 - 1.00) Based on data from 82,553 patients in 5 studies	74 per 1000	74 per 1000	High	Cold ischemic time has little or no difference on graft loss
		Difference: 0 more per 1000 (CI 95% 0 fewer - 0 more)			
HLA mismatch (1 mismatch increase)	Hazard Ratio 1.08 (CI 95% 1.07 - 1.09) Based on data from 171,446 patients in 4 studies	74 per 1000	80 per 1000	High	Increasing HLA mismatch slightly increases graft failure
		Difference: 6 more per 1000 (CI 95% 5 more - 6 more)			
Recipient Characteristics					
Age (10-year increase)	Hazard Ratio 1.16 (CI 95% 1.10 - 1.23) Based on data from 138,824 patients in 12 studies	74 per 1000	85 per 1000	High	Increasing recipient age slightly increases 1-year graft loss
		Difference: 11 more per 1000 (CI 95% 7 more - 16 more)			
Sex (male vs female)	Hazard Ratio 1.28 (CI 95% 0.98 - 1.67) Based on data from 176,972 patients in 9 studies	63 per 1000	81 per 1000	Low Due to seriuos inconsistency, Due to serious publication bias	Recipient sex may have little or no difference on graft loss
		Difference: 18 more per 1000 (CI 95% 2 fewer - 35 more)			
Race (white vs others)	Hazard Ratio 0.76 (CI 95% 0.52 - 1.11) Based on data from 169,596 patients in 2 studies	87 per 1000	66 per 1000	High	Recipient race has little or no difference on graft failure
		Difference: 21 fewer per 1000 (CI 95% 51 fewer - 8 more)			
BMI (1 kg/m2 increase)	Hazard Ratio 1.02 (CI 95% 0.99 - 1.04) Based on data from 51,881 patients in 4 studies	74 per 1000	75 per 1000	High	Recipient BMI has little or no impact on graft failure
		Difference: 1 more per 1000 (CI 95% 1 fewer - 3 more)			
Dialysis time (per 1-year increase)	Hazard Ratio 1.03 (CI 95% 1.02 - 1.03) Based on data from 51,776 patients in 3 studies	74 per 1000	76 per 1000	Moderate Due to serious risk of bias	Years on dialysis proabbly increases graft failure slightly
		Difference: 2 more per 1000 (CI 95% 1 fewer - 2 more)			
Diabetes (Yes vs no)	Hazard Ratio 0.99 (CI 95% 0.97 - 1.02) Based on data from 169,015 patients in 2 studies	74 per 1000	73 per 1000	High	Recipient diabetes has little or no difference on graft failure
		Difference: 1 more per 1000 (CI 95% 2 fewer - 1 more)			
Smoking (ever vs never)	Hazard Ratio 1.59 (CI 95% 1.34 - 1.90) Based on data from 3,156 patients in 2 studies	65 per 1000	104 per 1000	Moderate Due to serious imprecision	Pre-transplant recipient smoking probably increases graft failure slightly
		Difference: 39 more per 1000 (CI 95% 23 more - 55 more)			
Coronary Artery Disease (Yes vs no)	Hazard Ratio 1.15 (CI 95% 1.03 - 1.27) Based on data from 81,194 patients in 2 studies	73 per 1000	85 per 1000	Moderate Due to serious indirectness	Recipient coronary artery disease probably increases graft failure slightly
		Difference: 12 more per 1000 (CI 95% 2 more - 20 more)			
Hypertension (Yes vs no)	Hazard Ratio 0.99 (CI 95% 0.98 - 1.01) Based on data from 169,314 patients in 3 studies	74 per 1000	74 per 1000	High	Pre-transplant recipient hypertension has little or no difference on graft failure
		Difference: 1 fewer per 1000 (CI 95% 2 fewer - 0)			
Post-transplant Complications					
Delayed graft function (Yes vs no)	Hazard Ratio 1.89 (CI 95% 1.46 - 2.47) Based on data from 2,564 patients in 4 studies	63 per 1000	119 per 1000	Moderate Due to serious inconsistency	Delayed graft function probably increases graft failure
		Difference: 56 more per 1000 (CI 95% 31 more - 84 more)			
Acute rejection (Yes vs no)	Hazard Ratio 3.16 (CI 95% 1.86 - 5.38) Based on data from 48,768 patients in 7 studies	64 per 1000	203 per 1000	Moderate Due to serious inconsistency	Acute rejection probably increases graft failure
		Difference: 139 more per 1000 (CI 95% 60 more - 247 more)			

APPENDIX A – Characteristics and demographics of included studies

First Author	Year	Study Type	Multi-center vs Single-center	Name of Country or Registry	Inclusion Criteria	Recruitment Time Frame (Years)
Adekoya	2016	Retrospective	SC	UK	Deceased-donor transplants from donors ≥ 60 years old	1969 - 2009
An	2016	Retrospective	MC	Korea	Recipients ≥18 years old	1997-2012
Anderson	2015	Retrospective	MC	OPTN/UNOS	Recipients ≥18 years old with ECD kidneys	1987-2011
Andreoni	2013	Retrospective	MC	OPTN/UNOS	Recipients ≤55 years with no transplant history, primary graft from living or standard-criteria deceased donor	1987-2010
Andresdottir	2005	Retrospective	MC	Euro-transplant International foundation	Recipients ≥18 years of age with IgA nephropathy, who received a primary cadaveric renal graft	1990-2002
Andresdottir	2005	Retrospective	MC	Euro-transplant International foundation	Recipients ≥18 years of age with IgA nephropathy, who received a primary cadaveric renal graft	1990-2002
Asderakis	2001	Retrospective	SC	UK	First deceased donor kidney transplants	1990-1997
Bay	2013	Retrospective	MC	Denmark	All adult renal transplants	1998-2009
Boffa	2017	Retrospective	MC	UK	First adult kidney-only transplants	2003 - 2013
Brar	2013	Retrospective	MC	USRDS	First kidney transplant	2004 - 2009
Cardinal	2005	Retrospective	MC	Canada	Recipients ≥60 years with first deceased donor kidney transplant	1985 - 2000
Carrier	2012	Retrospective	MC	Canada	NR	2003 - 2009
Courtney	2007	Prospective	SC	Ireland	First deceased donor renal transplants	1986 - 2005
Diaz	2009	Retrospective	SC	Spain	NR	1980 - 2004
Dinis	2015	Prospective	SC	Portugal	Non-hyperimmunized recipients (PRA < 80%) submitted to first and single deceased renal transplantations	2006 - 2009
Faravardeh	2013	Retrospective	SC	USA	First adult kidney transplant recipients	1963 - 2012
Faravardeh	2013	Retrospective	SC	USA	First adult kidney transplant recipients	1963 - 2012
Faravardeh	2013	Retrospective	SC	USA	First adult kidney transplant recipients	1963 - 2012
Ferrer	2009	Retrospective	SC	Portugal	Deceased donor kidney transplant	2005 - 2009
Fuggle	2010	Retrospective	MC	UK	Adult recipients of first live donor kidney transplants	2000 - 2007
Grosso	2012	Retrospective	SC	Italy	Adult first kidney alone transplants	2000 - 2010
Heldal	2011	Retrospective	MC	Norway	All patients >70 years of age who received their first single kidney transplant	2000 - 2005
Heldal	2009	Retrospective	SC	Denmark	First kidney transplant	1990 - 2005
Heldal	2009	Retrospective	SC	Denmark	First kidney transplant	1990 - 2005
Heldal	2009	Retrospective	SC	Denmark	First kidney transplant	1990 - 2005
Huaman	2016	Retrospective	MC	OPTN/UNOS	All adults who underwent first, single organ deceased-donor kidney transplantation	2008 - 2013
Ilori	2015	Retrospective	MC	OPTN/UNOS	kidney transplant recipients aged ≥ 60 years	1996 - 2010
Kayler	2009	Retrospective	MC	USA	All kidney transplant recipients of adult donors	1995-2007
Koo	2015	Retrospective	SC	South Korea	NR	2000 - 2009
Kruger	2007	Retrospective	SC	Germany	First renal transplant	1995 - 2006
Lee	2014	Retrospective	SC	South Korea	First living donor kidney transplantations	2000 - 2011
Lin	2004	Retrospective	SC	Taiwan	First deceased donor transplants	1981-2000
Lynch	2009	Retrospective	SC	USA	All adult first kidney only transplants	2003-2008
Molnar	2012	Retrospective	MC	SRTR	All kidney transplant recipients	1998 - 2006
Molnar	2011	Retrospective	MC	USA	All kidney transplant recipients who underwent dialysis pre-transplant	2001-2006
Moore	2010	Retrospective	SC	UK	All consecutive kidney transplant	1996 - 2006
Nanmoku	2012	Prospective	SC	Japan	NR	2000 - 2009
Papalia	2010	Retrospective	SC	Italy	First time, adult, kidney-only transplant recipients	1998-2008
Nee	2013	Retrospective	MC	USRDS	First kidny transplantation patients with Lupus erythematosus	1995 - 2006
Redfield	2016	Retrospective	MC	OPTN/UNOS	Highly sensitized adult kidney transplant recipients with PRA ≥98%	1997-2014

SC – Single center; MC – multi-center; ECD – Extended Criteria Donor, PRA – Panel Reactive Antibodies; NR – Not Reported

APPENDIX A Continued– Characteristics and demographics of included studies

First Author	Year	Follow-up	Stratified Model	n cohort	n events	Definition of graft loss	Female Recipient	Female Donor
Adekoya	2016	NR	NA	112	41	RRT, Death Censored	38	59
An	2016	6.4 (0 - 17.8)***	NA	2902	286	RRT, nephrectomy, re-transplantation, Death Censored	1190	1239
Anderson	2015	NR	NA	25640	11691	NR	NR	NR
Andreoni	2013	6**	NA	168809	46854	NR, Death Censored	68705	78159
Andresdottir	2005	NR	Immunoglobulin A nephropathy	1207	93	NR, Death Censored	229	506
Andresdottir	2005	NR	Controls	7935	881	NR, Death Censored	4317	3142
Asderakis	2001	NR	NA	788	NR	NR, Death Included	254	NR
Bay	2013	NR	NA	676	97	NR, Death Censored	255	360
Boffa	2017	NR	NA	11,655	NR	NR, Death Included	NR	NR
Brar	2013	5.04*	NA	80,880	6,855	RRT	31637	NR
Cardinal	2005	6**	NA	256	124	RRT, re-transplant, Death Included	81	NR
Carrier	2012	3.5 (2)*	NA	1375	133	RRT, re-transplant, Death Included	715	359
Courtney	2007	8.2**	NA	707	198	RRT, Death Censored	331	352
Diaz	2009	8 (4.6)*	NA	250	65	NR, Death Included	87	NR
Dinis	2015	5.2*	NA	236	NR	NR	84	92
Faravardeh	2013	8.4*	Recipient age ≥65	364	NR	NR, Death Censored	136	NR
Faravardeh	2013	8.4*	Recipient age 50 - 64	1218	NR	NR, Death Censored	505	NR
Faravardeh	2013	8.4*	Recipient age <50	2900	NR	NR, Death Censored	1101	NR
Ferrer	2009	NR	NA	409	46	NR	140	NR
Fuggle	2010	3.5**	NA	3144	309	RRT, nephrectomy, Death Censored	1253	1737
Grosso	2012	13.6 (5.2)*	NA	376	52	NR	133	176
Heldal	2011	5.1 (0.1 - 9.7)***	NA	160	21	RRT, Death Censored	NR	NR
Heldal	2009	NR	Senior (60 - 69)	577	85	RRT, Death Censored	179	260
Heldal	2009	NR	Control (45 - 54)	563	98	RRT, Death Censored	197	282
Heldal	2009	NR	Elderly (70 - 81.5)	354	45	RRT, Death Censored	110	152
Huaman	2016	NR	NA	51048	NR	RRT, re-transplant, Death Included	20239	20585
Ilori	2015	4.3 (5.2)****	NA	44,013	6206	NR, Death Censored	16501	NR
Kayler	2009	NR	NA	99240	NR	NR	NR	18889
Koo	2015	NR	NA	709	65	RRT, nephrectomy, re-transplantation, Death Censored	321	317
Kruger	2007	3.65 (2.61)*	NA	352	NR	NR, Death Censored	112	NR
Lee	2014	NR	NA	201	15	NR, Death Censored	75	96
Lin	2004	5.61 (3.98)*	NA	299	162	RRT, Death Included	128	54
Lynch	2009	NR	NR	869	NR	NR	316	NR
Molnar	2012	3.9 (1.9 - 6.8)**	NA	145470	22876	RRT, re-transplant, Death Censored	57821	NR
Molnar	2011	6**	NR	8961	785	RRT, re-transplant, Death Censored	3316	NR
Moore	2010	5.75**	NA	697	301	RRT, re-transplant, Death Censored	262	285
Nanmoku	2012	NR	NA	564	NR	NR	210	346
Papalia	2010	5 (3.88)*	NA	206	NR	NR	71	NR
Nee	2013	5**	NA	4214	NR	NR, Death Included	60190	NR
Redfield	2016	NR	NA	107292	NR	NR	36284	51409

* - mean (standard-deviation)
** median (25th – 75th %)
*** median (range)
**** median (Interquartile range)

APPENDIX A Continued– Characteristics and demographics of included studies

First Author	Year	Recipient age	Donor age	Recipient Race: Black	Recipient Race: Caucasian	Recipient Race: Other	Recipient BMI	Donor BMI
Adekoya	2016	50.39 (13.72)*	64.71 (4)*	NR	NR	NR	NR	NR
An	2016	42 (33 - 51)**	40 (30 - 48)**	NR	NR	NR	22.1 (20.1 - 24.4)**	23.4 (21.2 - 25.6)**
Anderson	2015	57 (48 - 65)**	61 (55 - 65)**	736	13213	1904	26.7 (23.6 - 30.4)**	26.9 (23.8 - 30.8)**
Andreoni	2013	37.7 (12.6)*	34.4 (14)*	40363	95433	33013	NR	NR
Andresdottir	2005	NR	NR	NR	NR	NR	NR	NR
Andresdottir	2005	NR	NR	NR	NR	NR	NR	NR
Asderakis	2001	42.1 (15.5)*	NR	NR	NR	NR	NR	NR
Bay	2013	42.7 (14.7)*	47.3 (14.8)*	NR	NR	NR	NR	NR
Boffa	2017	18-29 (856), 30-39 (1580), 40-49 (2737), 50-59 (3116), 60+ (3366)	NR	807	9087	1761	NR	NR
Brar	2013	48.2*	38.8 (15)*	10127	67188	2777	27.1 (5.7)*	NR
Cardinal	2005	63 (61 - 65)**	42 (23 - 51)**	NR	NR	NR	25 (4)*	NR
Carrier	2012	51.3 (12.7)*	48.2 (11.5)*	NR	NR	NR	NR	NR
Courtney	2007	42 (16.7)*	37 (16.8)*	0	707	0	NR	NR
Diaz	2009	47.7 (14.2)*	NR	NR	NR	NR	NR	NR
Dinis	2015	49.35 (14)*	46.65 (16.5)*	NR	NR	NR	NR	NR
Faravardeh	2013	69.5 (3.5)*	NR	12	328	24	NR	NR
Faravardeh	2013	57 (4.2)*	NR	49	1095	73	NR	NR
Faravardeh	2013	35.4 (8.7)*	NR	119	2630	146	NR	NR
Ferrer	2009	46.4 (10.3*)	44.3 (13.8)*	NR	NR	NR	NR	NR
Fuggle	2010	39.8 (29.2 - 49.2)**	47.3 (38.6 - 55.3)**	110	2762	278	NR	NR
Grosso	2012	48.1 (12.25)*	51.2 (18.1)*	NR	NR	NR	NR	NR
Heldal	2011	73.6 (70 - 81.1)***	55.2 (4 - 82)***	NR	NR	NR	NR	NR
Heldal	2009	64.8 (62.4 - 67.4)**	51.7 (1 - 83)***	1 - 83	NR	NR	NR	NR
Heldal	2009	50.2 (47.6 - 52.3)**	47.1 (0 - 80)***	0 - 80	NR	NR	NR	NR
Heldal	2009	73.5 (71.3 - 75.8)**	51.3 (2 - 82)***	2 - 82	NR	NR	NR	NR
Huaman	2016	53.8 (12.9)*	37.4 (16.6)*	7186	NR	NR	28.4 (5.4)*	27.4 (6.8)*
Ilori	2015	65 (7)*	43.2 (16)*	8903	27481	7629	NR	NR
Kayler	2009	NR	38 **	112%	NR	NR	NR	NR
Koo	2015	41.4 (33.2 - 49.4)**	40.5 (31.3 - 47.8)**	NR	NR	NR	22.2 (20 - 25)**	NR
Kruger	2007	50.5 (13.8)*	50.3 (15.8)*	NR	NR	NR	NR	NR
Lee	2014	40.02 (10)*	38.5 (10.5)*	NR	NR	NR	22.5 (2.8)*	23.1 (2.9)*
Lin	2004	35 (10)*	29.5 (11.01)*	NR	NR	NR	NR	NR
Lynch	2009	50.01 **	NR	149	NR	NR	NR	NR
Molnar	2012	47.1 (4.6)*	38.5 (14.4)*	33276	85827	26367	26.7 (5.4)*	NR
Molnar	2011	48 (13)*	39 (15)*	2420	NR	NR	26.6 (5.7)*	NR
Moore	2010	41.6 (16.7)*	36.5 (16.8)*	0	697	0	NR	NR
Nanmoku	2012	40.3 (12.35)*	53.47 (10.85)*	NR	NR	NR	NR	NR
Papalia	2010	43.94 (12.69)*	46.59 (8.96)*	0	206	0	24.3 (2.83)*	NR
Nee	2013	46.1 (15.2)*	37.7 (14.9)*	34720	15398	15335	NR	NR
Redfield	2016	50.24 (13.6)*	39.53 (15.17)*	24773	60158	22334	27.54 (5.52)*	26.79 (5.64)*

APPENDIX A Continued– Characteristics and demographics of included studies

First Author	Year	Ischemic Time	Acute Rejection	Dialysis pre-transplant	Duration of Dialysis (years) (mean/median)	ECD	Deceased donor	DGF	Diabetes mellitus	Hypertension
Adekoya	2016	17.1 (6.8)*	43	NR	NR	NR	112	45	NR	36
An	2016	NR	NR	NR	NR	NR	642	NR	481	2441
Anderson	2015	19 (14 - 26)**	NR	20985	NR	25640	25640	8421	NR	16910
Andreoni	2013	NR	NR	NR	NR	0	0	NR	2104	8784
Andresdottir	2005	NR	NR	NR	NR	NR	NR	NR	NR	NR
Andresdottir	2005	NR	NR	NR	NR	NR	NR	NR	NR	NR
Asderakis	2001	25.26 (7.52)*	707	703	NR	NR	NR	184	NR	NR
Bay	2013	13.9 (9.47)*	153	NR	NR	NR	456	NR	86	62
Boffa	2017	13.04 (11.12)*	NR	NR	NR	NR	NR	NR	NR	NR
Brar	2013	13.04 (11.12)*	NR	57577	NR	NR	50261	13266	24013	62909
Cardinal	2005	18 (7.3)*	NR	NR	2	NR	256	NR	37	NR
Carrier	2012	Nr	NR	NR	NR	282	792	NR	NR	NR
Courtney	2007	22.3 (18.8 - 26.7)**	NR	NR	NR	NR	707	NR	NR	NR
Diaz	2009	17.2 (5.01)*	NR	NR	NR	NR	NR	NR	17	NR
Dinis	2015	17.2 (5.01)*	20	NR	NR	NR	236	57	NR	NR
Faravardeh	2013	NR	56	NR	2.22	51	NR	NR	101	NR
Faravardeh	2013	NR	201	NR	2.08	70	NR	NR	369	NR
Faravardeh	2013	16.1 (5.97)*	490	NR	1.4	57	NR	NR	691	NR
Ferrer	2009	18.3 (5.59)*	86	NR	4.4	100	NR	90	NR	NR
Fuggle	2010	NR	NR	NR	NR	NR	0	NR	NR	NR
Grosso	2012	16.1 (5.97)*	NR	NR	4.2	NR	376	90	NR	NR
Heldal	2011	12 (1 - 29)***	36	144	1.4	NR	NR	48	22	NR
Heldal	2009	11 (1 - 32)***	282	560	0.6	NR	321	51	NR	NR
Heldal	2009	7 (1 - 35)***	142	538	1.2	NR	304	85	NR	NR
Heldal	2009	13 (1 - 28)***	289.00	344	0.8	NR	404	86	NR	NR
Huaman	2016	17.7 (17.1)*	2982	NR	3.72	NR	51048	13006	19321	NR
Ilori	2015	15 (18)****	4436	37848	2.47	9412	30799	NR	NR	NR
Kayler	2009	NR	NR	NR	NR	NR	256	NR	NR	NR
Koo	2015	1.1 (0.8 - 2.3)**	198	NR	1.2	NR	153	35	99	NR
Kruger	2007	12 (7.6)*	124	329	4	NR	272	60	60	336
Lee	2014	0.9 (0.35)*	NR	121	14.7	NR	0	NR	NR	NR
Lin	2004	NR	78	NR	NR	NR	299	NR	8	118
Lynch	2009	NR	NR	NR	NR	NR	446	NR	299	NR
Molnar	2012	12.3 (1.7 - 20.6)**	7306	NR	NR	22516	80084	NR	28853	119869
Molnar	2011	14.4 (10.6)*	NR	NR	NR	1703	NR	1951	2420	NR
Moore	2010	NR	NR	NR	NR	NR	697	NR	NR	NR
Nanmoku	2012	NR	NR	468	5.4	NR	51	NR	NR	NR
Papalia	2010	11.04 (6.83)*	37	NR	5.9	NR	172	59	31	131
Nee	2013	NR	NR	105606	2.3	49274	93780	23323	30474	NR
Redfield	2016	NR	NR	76565	NR	68394	12437	NR	NR	NR

APPENDIX B – Risk of bias of included studies

First Author	Year	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting	Overall
Adekoya	2016	Low	NR	Moderate	Low	Low	Low	Low
An	2016	Low	NR	Low	Low	Low	Low	Low
Anderson	2015	Low	Low	Moderate	Moderate	Low	Low	High
Andreoni	2013	Low	Low	Low	Moderate	Low	Low	Low
Andresdottir	2005	Low	Low	Moderate	Moderate	High	High	High
Andresdottir	2005	Low	Low	Moderate	Moderate	High	High	High
Asderakis	2001	Low	Moderate	Low	Low	Low	Low	Low
Bay	2013	Low	Low	Low	Moderate	Low	Low	Low
Boffa	2017	Low	Low	Low	Low	Low	High	High
Brar	2013	Moderate	NR	Low	Low	Low	Low	Low
Cardinal	2005	Low	Low	Low	Low	Low	Low	Low
Carrier	2012	High	NR	Low	Low	Low	High	High
Courtney	2007	Low	NR	Low	Low	High	High	High
Diaz	2009	Low	NR	High	Low	Moderate	Low	High
Dinis	2015	Low	NR	Moderate	Moderate	Moderate	Low	High
Faravardeh	2013	Low	NR	Low	Moderate	Low	Low	Low
Faravardeh	2013	Low	NR	Low	Moderate	Low	Low	Low
Faravardeh	2013	Low	NR	Low	Moderate	Low	Low	Low
Ferrer	2009	Moderate	NR	Low	Moderate	Low	High	High
Fuggle	2010	Moderate	NR	Moderate	Low	Low	High	High
Grosso	2012	Low	NR	Low	Moderate	Low	High	High
Heldal	2011	Low	NR	Low	Low	Moderate	High	High
Heldal	2009	Low	NR	Low	Low	Moderate	High	High
Heldal	2009	Low	NR	Low	Low	Moderate	High	High
Heldal	2009	Low	NR	Low	Low	Moderate	High	High
Huaman	2016	Low	NR	Moderate	Low	Moderate	Low	High
Ilori	2015	Low	NR	Low	Moderate	Low	Low	Low
Kayler	2009	Moderate	NR	Moderate	Moderate	Low	Low	High
Koo	2015	Moderate	NR	Moderate	Low	Low	Low	Low
Kruger	2007	Low	NR	Low	High	Moderate	Moderate	High
Lee	2014	Low	NR	Moderate	Low	Low	High	High
Lin	2004	Low	NR	Low	Low	Low	Low	Low
Lynch	2009	Low	NR	Low	Moderate	Low	Moderate	High
Molnar	2012	Moderate	NR	Low	Low	Low	High	High
Molnar	2011	Low	NR	Low	Low	Low	Low	Low
Moore	2010	Low	Low	Low	Moderate	Low	Low	Low
Nanmoku	2012	Low	NR	High	Low	Moderate	Low	High
Papalia	2010	Low	NR	Low	Low	Low	Low	Low
Nee	2013	Low	Low	Low	Low	Low	Low	Low
Redfield	2016	Low	NR	Low	Low	Low	Moderate	Low

NR – not reported

APPENDIX C – Systematic Search Strategy

All searches executed on February 6, 2017:

Database: Ovid MEDLINE(R) <1946 to January Week 4 2017>
Search Strategy:
---

1 [kidney transplantation] (0)
2 Kidney Transplantation/ (85861)
3 exp kidney/tr (69)
4 (kidney? adj2 transplant*).mp. (90052)
5 (kidney? adj2 graft*).mp. (3806)
6 (kidney? adj2 allograft*).mp. (3248)
7 (kidney? adj2 allotransplant*).mp. (213)
8 (kidney? adj2 heterograft*).mp. (2)
9 (kidney? adj2 heterotransplant*).mp. (1)
10 (kidney? adj2 homotransplant*).mp. (109)
11 (kidney? adj2 homograft*).mp. (47)
12 (renal adj2 transplant*).mp. (40446)
13 (renal adj2 graft*).mp. (3147)
14 (renal adj2 allograft*).mp. (11163)
15 (renal adj2 allotransplant*).mp. (428)
16 (renal adj2 heterograft*).mp. (3)
17 (renal adj2 heterotransplant*).mp. (19)
18 (renal adj2 homotransplant*).mp. (330)
19 (renal adj2 homograft*).mp. (249)
20 or/2-19 (96397)
21 [graft loss] (0)
22 graft rejection/ (54629)
23 (graft? adj2 loss*).mp. (5929)
24 (transplant adj2 loss*).mp. (210)
25 or/22-24 (58046)
26 20 and 25 (23756)
27 [predictors/risk/prognostic factors] (0)
28 predict*.mp. (1089664)
29 scor*.tw. (611522)
30 observ*.mp. (2588936)
31 validat*.mp. (342123)
32 exp risk/ (985946)
33 risk*.mp. (1848735)
34 exp Cohort Studies/ (1609759)
35 between group*.tw. (83478)
36 exp prognosis/ (1318869)

37 (prognos* adj2 factor*).mp. (72868)
38 (prognos* adj2 value*).mp. (35580)
39 (associat* adj2 factor*).mp. (117849)
40 (independent adj2 factor*).mp. (57112)
41 (multivariate adj2 factor*).mp. (3082)
42 (multivariable* adj2 factor*).mp. (504)
43 exp Regression Analysis/ (351660)
44 regression*.mp. (535453)
45 (hazard* adj2 model*).mp. (73870)
46 (cox adj2 model*).mp. (14261)
47 (hazard* adj2 ratio*).mp. (65084)
48 exp survival analysis/ (229756)
49 or/28-48 (6848630)
50 26 and 49 (13920)
51 animals/ not (animals/ and humans/) (4277235)
52 50 not 51 (13507)
53 limit 52 to ("all infant (birth to 23 months)" or "all child (0 to 18 years)") (3110)
54 limit 52 to "all adult (19 plus years)" (8531)
55 52 not 53 (10397)
56 54 or 55 (12725)
57 limit 56 to yr="2000 -Current" (8622)
58 from 57 keep 1-4000 (4000)
59 from 57 keep 4001-8622 (4622)
60 remove duplicates from 58 (3862)
61 remove duplicates from 59 (4620)
62 60 or 61 (8482)

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <February 03, 2017>
Search Strategy:
---
1 (kidney? adj2 transplant*).mp. (3048)
2 (kidney? adj2 graft*).mp. (199)
3 (kidney? adj2 allograft*).mp. (212)
4 (kidney? adj2 allotransplant*).mp. (3)
5 (kidney? adj2 heterograft*).mp. (0)
6 (kidney? adj2 heterotransplant*).mp. (2)
7 (kidney? adj2 homotransplant*).mp. (0)
8 (kidney? adj2 homograft*).mp. (0)
9 (renal adj2 transplant*).mp. (2523)
10 (renal adj2 graft*).mp. (185)
11 (renal adj2 allograft*).mp. (464)
12 (renal adj2 allotransplant*).mp. (6)
13 (renal adj2 heterograft*).mp. (0)
14 (renal adj2 heterotransplant*).mp. (0)

15 (renal adj2 homotransplant*).mp. (0)
16 (renal adj2 homograft*).mp. (1)
17 (graft? adj2 loss*).mp. (517)
18 (transplant adj2 loss*).mp. (20)
19 or/1-16 (4867)
20 17 or 18 (535)
21 19 and 20 (337)
22 limit 21 to yr="2000 -Current" (325)

Database: Embase <1974 to 2017 February 03>
Search Strategy:
---
1 [kidney transplantation] (0)
2 exp kidney transplantation/ (136804)
3 (kidney? adj2 transplant*).mp. (127620)
4 (kidney? adj2 graft*).mp. (52987)
5 (kidney? adj2 allograft*).mp. (16323)
6 (kidney? adj2 allotransplant*).mp. (277)
7 (kidney? adj2 heterograft*).mp. (3)
8 (kidney? adj2 heterotransplant*).mp. (2)
9 (kidney? adj2 homotransplant*).mp. (71)
10 (kidney? adj2 homograft*).mp. (33)
11 (renal adj2 transplant*).mp. (62067)
12 (renal adj2 graft*).mp. (6493)
13 (renal adj2 allograft*).mp. (15758)
14 (renal adj2 allotransplant*).mp. (503)
15 (renal adj2 heterograft*).mp. (3)
16 (renal adj2 heterotransplant*).mp. (12)
17 (renal adj2 homotransplant*).mp. (296)
18 (renal adj2 homograft*).mp. (210)
19 or/2-18 (148461)
20 [graft loss] (0)
21 graft failure/ (30676)
22 kidney allograft rejection/ (3283)
23 kidney graft rejection/ (14325)
24 (graft? adj2 loss*).mp. (12046)
25 (transplant adj2 loss*).mp. (422)
26 or/21-25 (48978)
27 19 and 26 (27891)
28 [predictors/risk/prognostic factors] (0)
29 predict*.tw. (1566846)
30 exp methodology/ (5010109)
31 validat*.tw. (515387)

32 risk*.mp. (3024807)
33 exp epidemiology/ (2902080)
34 prognosis/ (590179)
35 prognostic assessment/ (3012)
36 (prognos* adj2 factor*).mp. (123954)
37 (prognos* adj2 value*).mp. (60036)
38 (associat* adj2 factor*).mp. (185971)
39 (independent adj2 factor*).mp. (96859)
40 (multivariate adj2 factor*).mp. (5218)
41 (multivariable* adj2 factor*).mp. (937)
42 exp regression analysis/ (532262)
43 regression*.mp. (841942)
44 (hazard* adj2 model*).mp. (123772)
45 (cox adj2 model*).mp. (28915)
46 (hazard* adj2 ratio*).mp. (113253)
47 survival analysis/ (3038)
48 (survival adj2 analy*).mp. (55752)
49 or/29-48 (9821988)
50 27 and 49 (17324)
51 (exp animals/ or exp animal experimentation/ or nonhuman/) not ((exp animals/ or exp animal experimentation/ or nonhuman/) and exp human/) (5926819)
52 50 not 51 (17120)
53 limit 52 to (embryo <first trimester> or infant <to one year> or child <unspecified age> or preschool child <1 to 6 years> or school child <7 to 12 years> or adolescent <13 to 17 years>) (2277)
54 limit 52 to (adult <18 to 64 years> or aged <65+ years>) (8833)
55 52 not 53 (14843)
56 54 or 55 (16327)
57 limit 56 to (book or book series or chapter or conference abstract or conference paper or conference proceeding or "conference review") (5242)
58 56 not 57 (11085)
59 limit 58 to yr="2000 -Current" (9957)
60 from 59 keep 1-5000 (5000)
61 from 59 keep 5001-9957 (4957)
62 remove duplicates from 60 (4582)
63 remove duplicates from 61 (4940)
64 62 or 63 (9522)

Database: EBM Reviews - Cochrane Central Register of Controlled Trials <November 2016>
Search Strategy:
---
1 [kidney transplantation] (0)
2 Kidney Transplantation/ (3195)
3 exp kidney/tr (0)

4 (kidney? adj2 transplant*).mp. (6017)
5 (kidney? adj2 graft*).mp. (1489)
6 (kidney? adj2 allograft*).mp. (364)
7 (kidney? adj2 allotransplant*).mp. (5)
8 (kidney? adj2 heterograft*).mp. (0)
9 (kidney? adj2 heterotransplant*).mp. (0)
10 (kidney? adj2 homotransplant*).mp. (2)
11 (kidney? adj2 homograft*).mp. (2)
12 (renal adj2 transplant*).mp. (3780)
13 (renal adj2 graft*).mp. (324)
14 (renal adj2 allograft*).mp. (921)
15 (renal adj2 allotransplant*).mp. (21)
16 (renal adj2 heterograft*).mp. (0)
17 (renal adj2 heterotransplant*).mp. (0)
18 (renal adj2 homotransplant*).mp. (3)
19 (renal adj2 homograft*).mp. (0)
20 or/2-19 (6860)
21 [graft loss] (0)
22 graft rejection/ (2010)
23 (graft? adj2 loss*).mp. (753)
24 (transplant adj2 loss*).mp. (35)
25 or/22-24 (2566)
26 20 and 25 (1799)
27 [predictors/risk/prognostic factors] (0)
28 predict*.mp. (55289)
29 scor*.tw. (117562)
30 observ*.mp. (155803)
31 validat*.mp. (15583)
32 exp risk/ (30591)
33 risk*.mp. (116134)
34 exp Cohort Studies/ (123483)
35 between group*.tw. (412183)
36 exp prognosis/ (118499)
37 (prognos* adj2 factor*).mp. (4223)
38 (prognos* adj2 value*).mp. (1824)
39 (associat* adj2 factor*).mp. (5529)
40 (independent adj2 factor*).mp. (2939)
41 (multivariate adj2 factor*).mp. (412)
42 (multivariable* adj2 factor*).mp. (62)
43 exp Regression Analysis/ (16965)
44 regression*.mp. (33916)
45 (hazard* adj2 model*).mp. (7152)
46 (cox adj2 model*).mp. (1654)
47 (hazard* adj2 ratio*).mp. (11573)

48 exp survival analysis/ (16426)
49 or/28-48 (596513)
50 26 and 49 (1578)
51 limit 50 to yr="2000 -Current" (1048)

Database: EBM Reviews - Cochrane Database of Systematic Reviews <2005 to February 02, 2017>
Search Strategy:
---
1 (kidney? adj2 transplant*).mp. (164)
2 (kidney? adj2 graft*).mp. (20)
3 (kidney? adj2 allograft*).mp. (15)
4 (kidney? adj2 allotransplant*).mp. (0)
5 (kidney? adj2 heterograft*).mp. (0)
6 (kidney? adj2 heterotransplant*).mp. (0)
7 (kidney? adj2 homotransplant*).mp. (0)
8 (kidney? adj2 homograft*).mp. (0)
9 (renal adj2 transplant*).mp. (78)
10 (renal adj2 graft*).mp. (4)
11 (renal adj2 allograft*).mp. (6)
12 (renal adj2 allotransplant*).mp. (0)
13 (renal adj2 heterograft*).mp. (0)
14 (renal adj2 heterotransplant*).mp. (0)
15 (renal adj2 homotransplant*).mp. (1)
16 (renal adj2 homograft*).mp. (0)
17 (graft? adj2 loss*).mp. (61)
18 (transplant adj2 loss*).mp. (6)
19 or/1-16 (229)
20 17 or 18 (62)
21 19 and 20 (43)

PubMed Query	Items found
Search (((((("kidney"[All Fields] OR "kidneys"[All Fields] OR "renal"[All Fields])) AND ("transplant"[Text Word] OR "transplants"[Text Word] OR "transplantation"[Text Word] OR "transplanting"[Text Word] OR "transplanted"[Text Word] OR graft*[Text Word] OR allograft*[Text Word] OR allotransplant*[Text Word] OR heterograft*[Text Word] OR heterotransplant*[Text Word] OR homotransplant*[Text Word] OR homograft*[Text Word]))))) AND ((("Graft"[Text Word] OR "grafts"[Text Word] OR "transplant"[Text Word])) AND loss*[Text Word]))) AND ((((((((((predict*[Title/Abstract] OR scor*[Title/Abstract] OR observ*[Title/Abstract] OR "validation"[Title/Abstract] OR	259

PubMed Query	Items found
"validate"[Title/Abstract] OR risk*[Title/Abstract] OR "cohort"[Title/Abstract])) OR group*[Text Word]) OR regression*[Text Word]) OR (hazard*[Text Word] AND model*[Text Word])) OR ("cox"[Text Word] AND model*[Text Word])) OR (hazard*[Text Word] AND ratio*[Text Word]))) OR (((prognos*[Text Word] OR associat*[Text Word] OR "independent"[Text Word] OR "multivariate"[Text Word] OR multivariable*[Text Word])) AND factor*[Text Word]))) AND ((pubstatusaheadofprint OR publisher[sb] OR pubmednotmedline[sb]))) **Filters:** Publication date from 2000/01/01 to 2017/12/31	

APPENDIX D – Supplemental Methods

Study selection

Among observational studies of adult (≥18 years) kidney recipients receiving their first transplant, we included studies evaluating the association between any prognostic factors and 1-year graft-loss using multivariable analysis (i.e., Cox proportional hazards or logistic regression models). Reasoning that donor age is widely perceived as an important predictor of graft loss, we excluded studies that did not adjust for donor age in their final multivariable model. We also excluded studies with <20 events in the follow-up period, and those that assessed graft loss beyond the first year if they did not use a time to event analysis that followed the proportional hazard assumption. In order to capture more contemporary transplant cohorts, we excluded studies published before 2000. We did not restrict by language or publication status. If two studies drew their sample from the same population and reported on the same predictors, we included the study with larger sample size. For this reason, due to mandatory reporting from US centres to the United Network of Organ Sharing (UNOS) registry, we only considered US studies using this registry, and excluded US studies from individual centres if the risk factors of interest were reported in the UNOS registry.

Working in pairs, reviewers independently screened titles and abstracts of identified citations and evaluated the full text of articles deemed potentially eligible by either reviewer. We resolved disagreements through discussion or adjudication by a third reviewer.

Data abstraction

Three pairs of independent reviewers, working in duplicate, extracted data from eligible studies. We extracted data related to center(s) at which data were collected, time frame of recruitment, and the definition and number of graft failures. Reviewers recorded the general

characteristics of cohorts from the individual studies: recipient age, sex, body mass index (BMI), race, diabetes, hypertension, duration of time on dialysis, immunosuppressive regiment, delayed graft function, acute rejection, donor age, sex, type (live vs deceased), quality (standard vs extended criteria), ischemic time, and sex mismatch between donor and recipient.

From each eligible study, reviewers abstracted key information about each predictor: the definition, estimate of effect, and confidence interval. We also recorded information about the outcome definitions; for example, whether graft loss was death-censored or not.

Data Synthesis

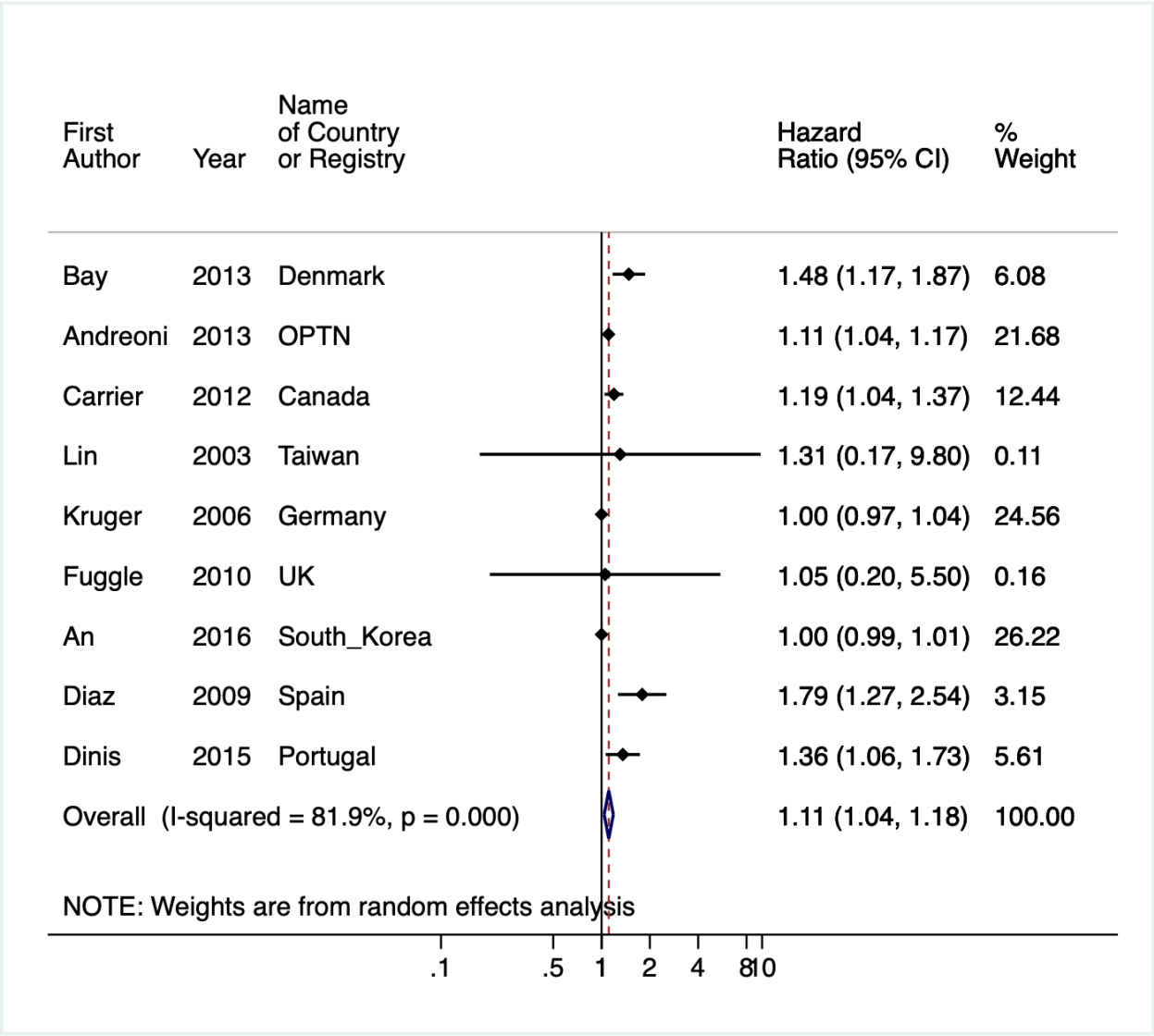
When studies reported continuous prognostic variables as ordinal variables, if we observed a linear relation between the increasing order and risk of graft loss, we averaged the beta-coefficients across categories to obtain the estimate of effect associated with a unit change for the meta-analysis. If we observed no linear relation, we reported the effect estimate separately for each category. For studies that had stratified the overall cohort into two or more groups and conducted the same regression model separately in each group, we pooled the effect estimates for the predictors to obtain one for the entire cohort. We pooled HR estimates for each predictor through inverse variance analysis using random-effects meta-analysis. If only two studies assessed the same predictor with one much larger than the other, we pooled results using the fixed effect framework.

Subgroup hypotheses

Risk of bias: We hypothesized greater strength of association between the prognostic factor and the outcome of interest for studies classified as high risk of bias compared to low risk of bias. Definition of outcome: The primary analysis for each predictor in this review pooled those

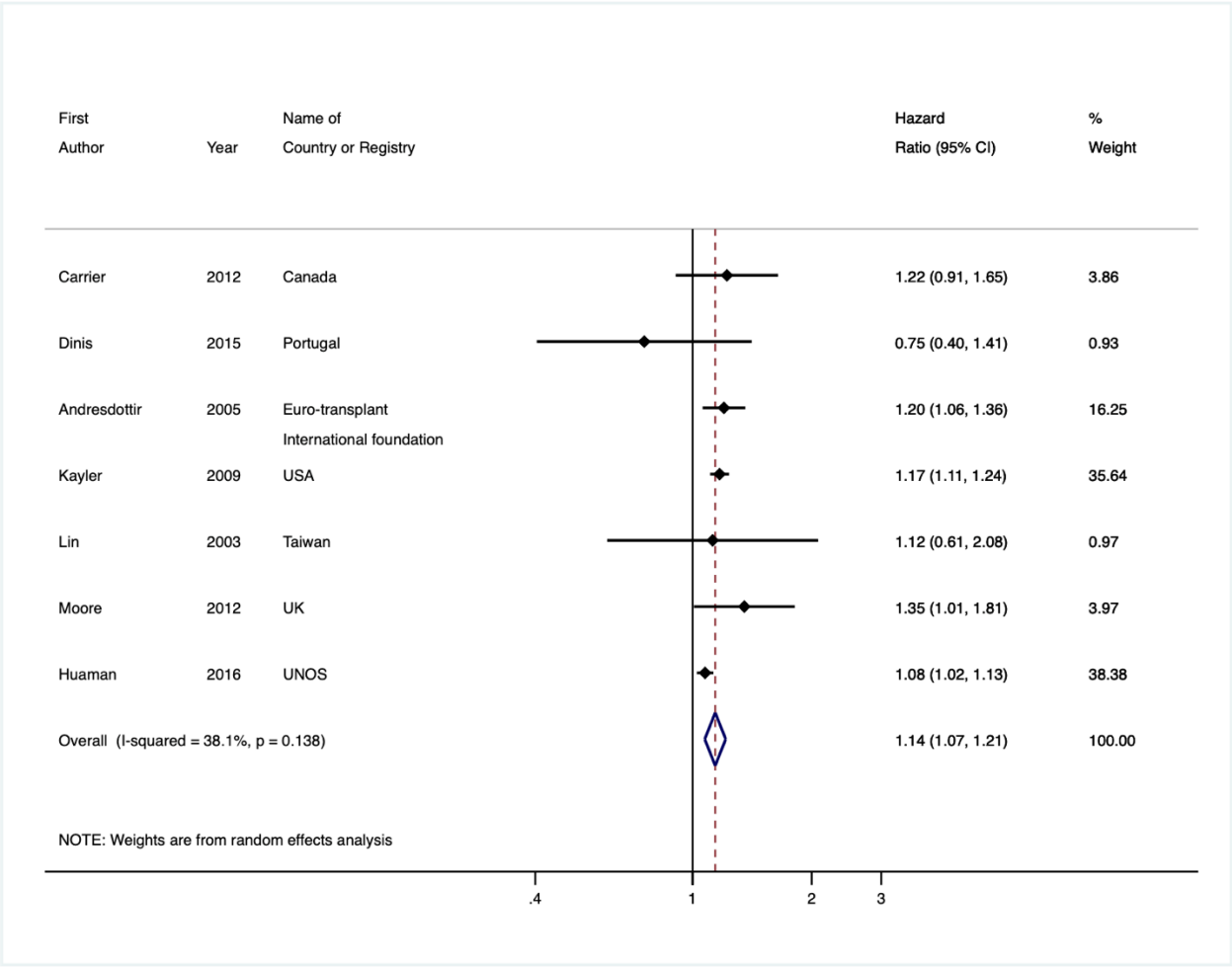
studies reporting on death-censored graft-loss at one year with those studies considering a non-death-censored graft loss. We postulated that the power of predictor variables may differ depending on which of these two outcomes measures the studies used. Moreover, we postulated that the direction of this difference may vary for donor-related, transplant related or recipient-related risk factors. For donor-specific and transplant-specific factors, we postulated a stronger association with death-censored graft-failure compared to the composite of graft failure and all-cause mortality. For recipient-specific factors, we postulated the opposite: a stronger relation with graft failure combined with all-cause mortality.

APPENDIX E – Meta-analysis of donor age



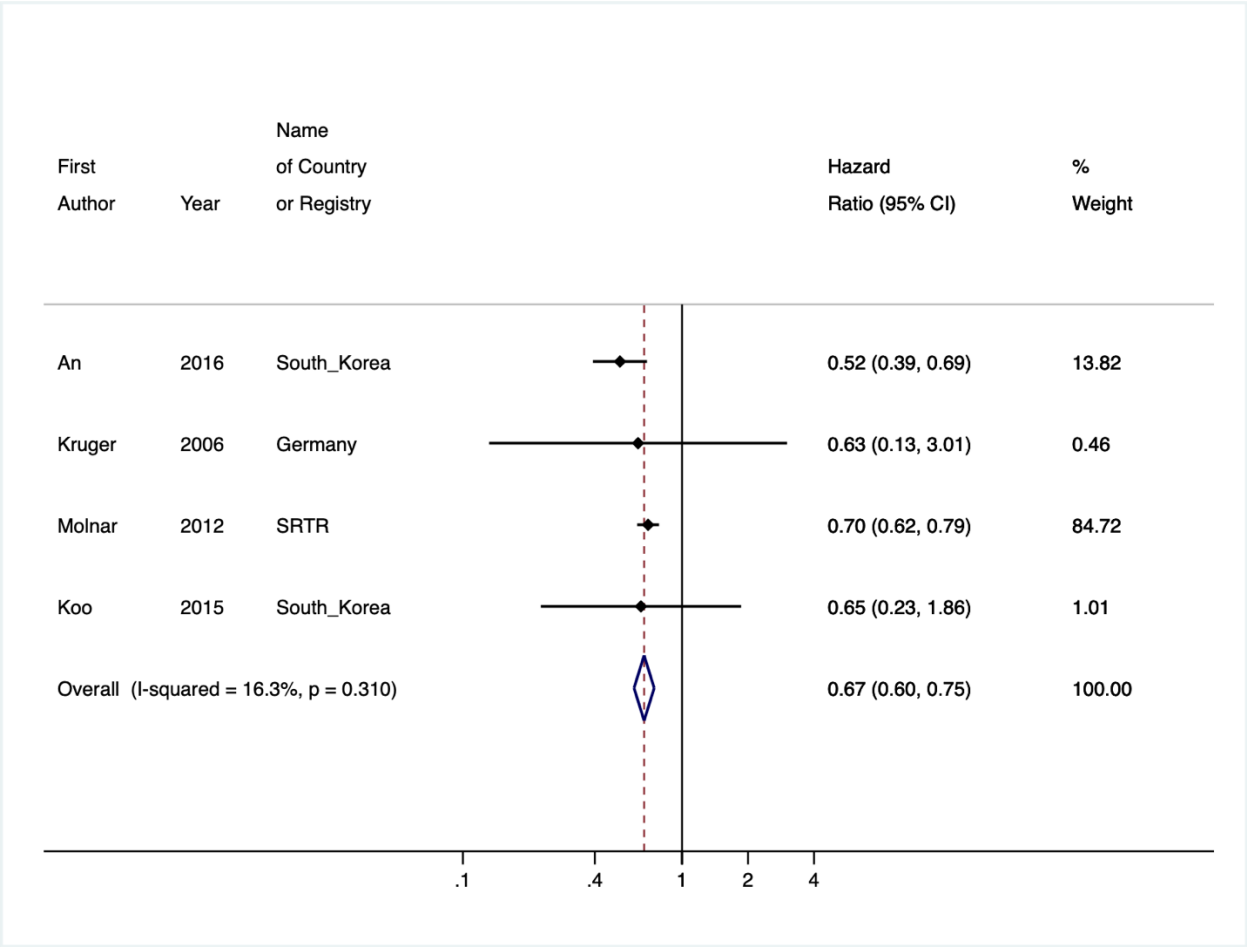
HR for every 10-year increase in age

APPENDIX F – Meta-analysis of donor sex



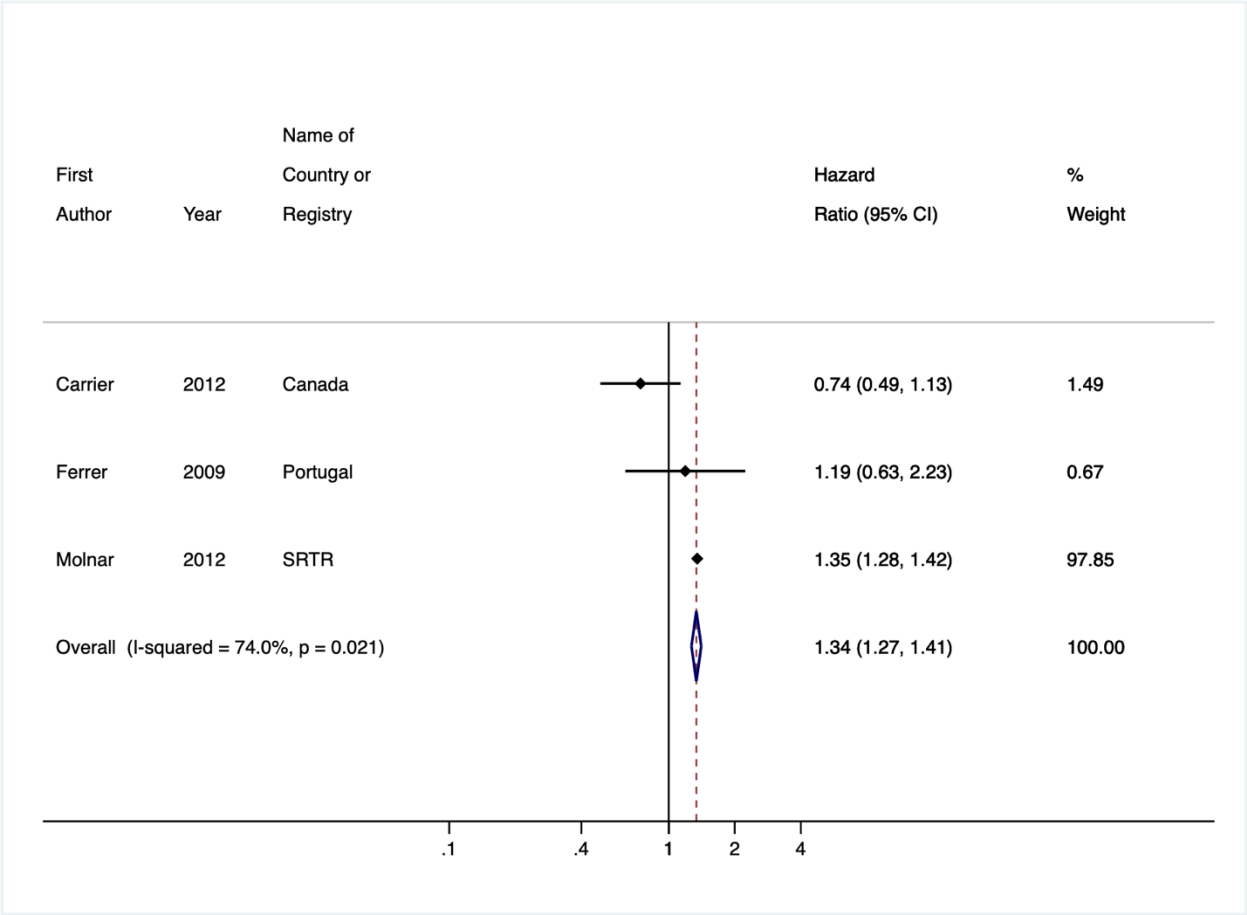
HR for female sex compared to male sex

APPENDIX G – Meta-analysis of donor type



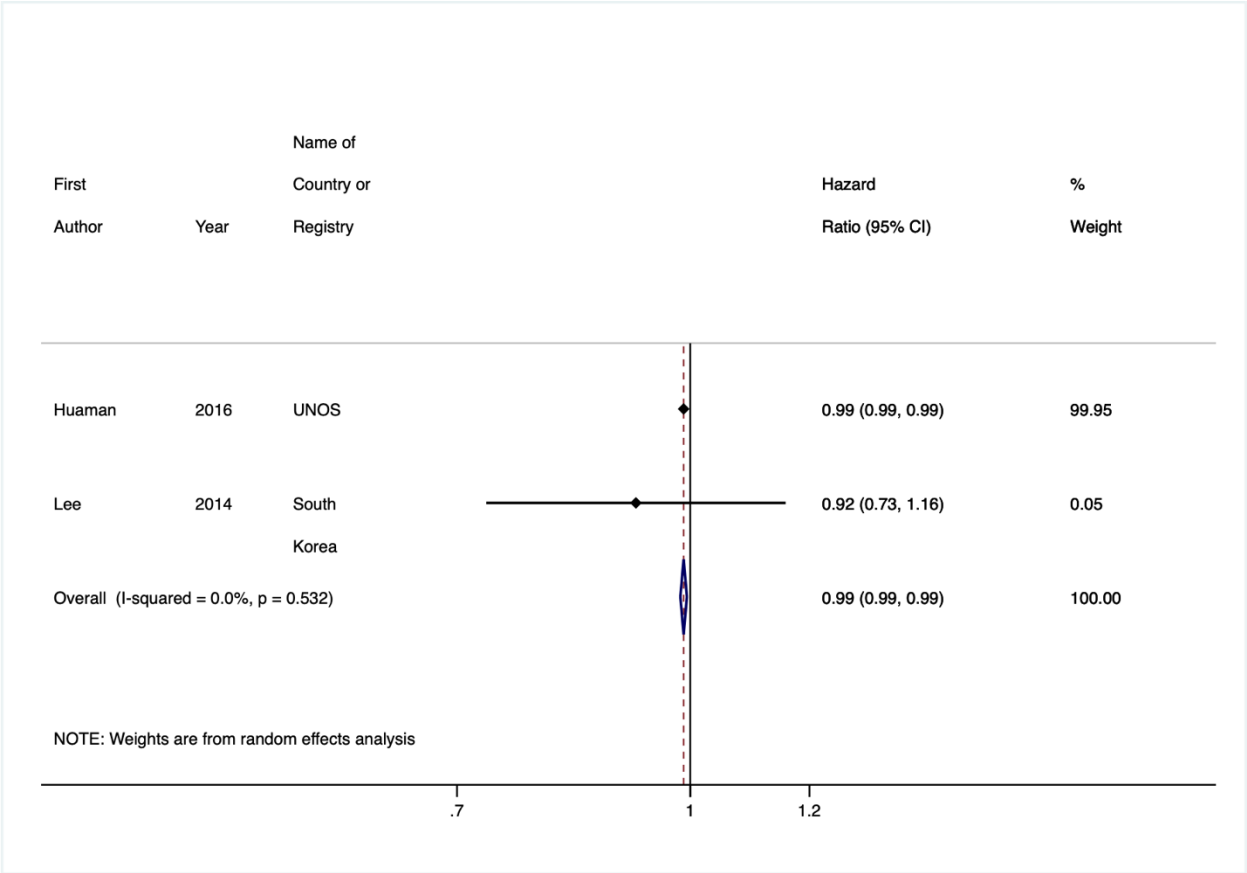
HR for living donors compared to deceased donors

APPENDIX H – Meta-analysis of donor quality



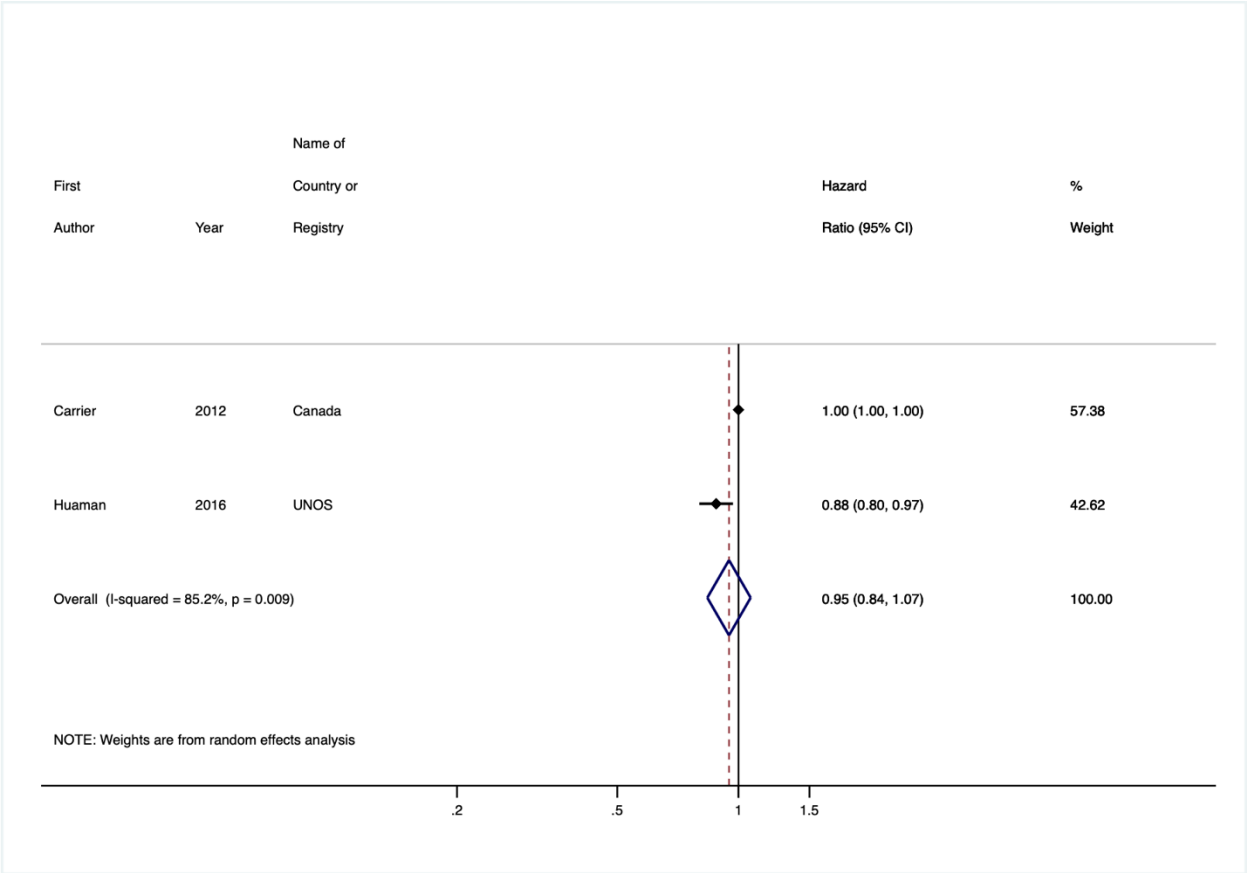
HR for extended criteria donors compared to normal criteria donors

APPENDIX I – Meta-analysis of donor body mass index



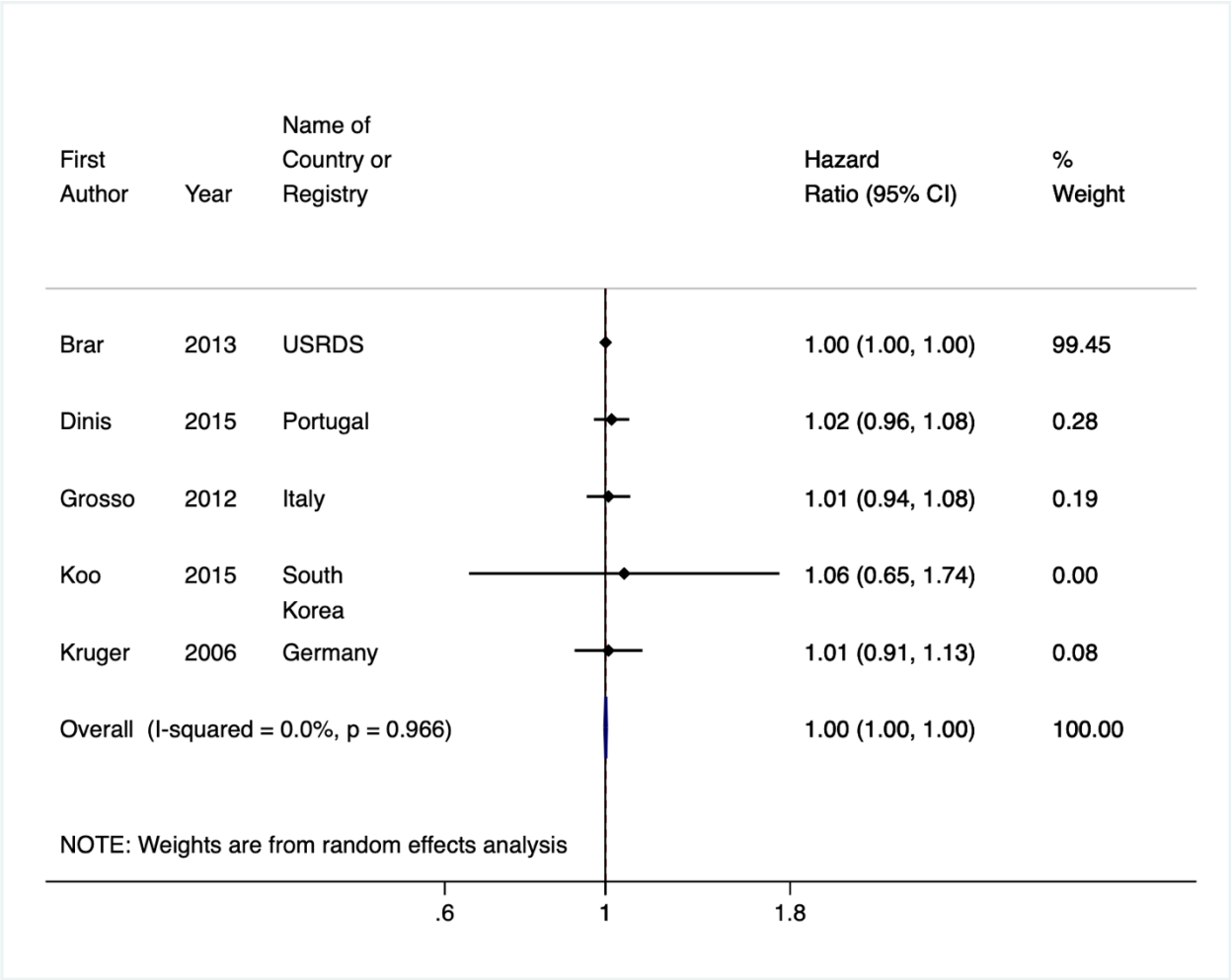
HR for every 1 kg/m^2 increase

APPENDIX J – Meta-analysis of donor creatinine



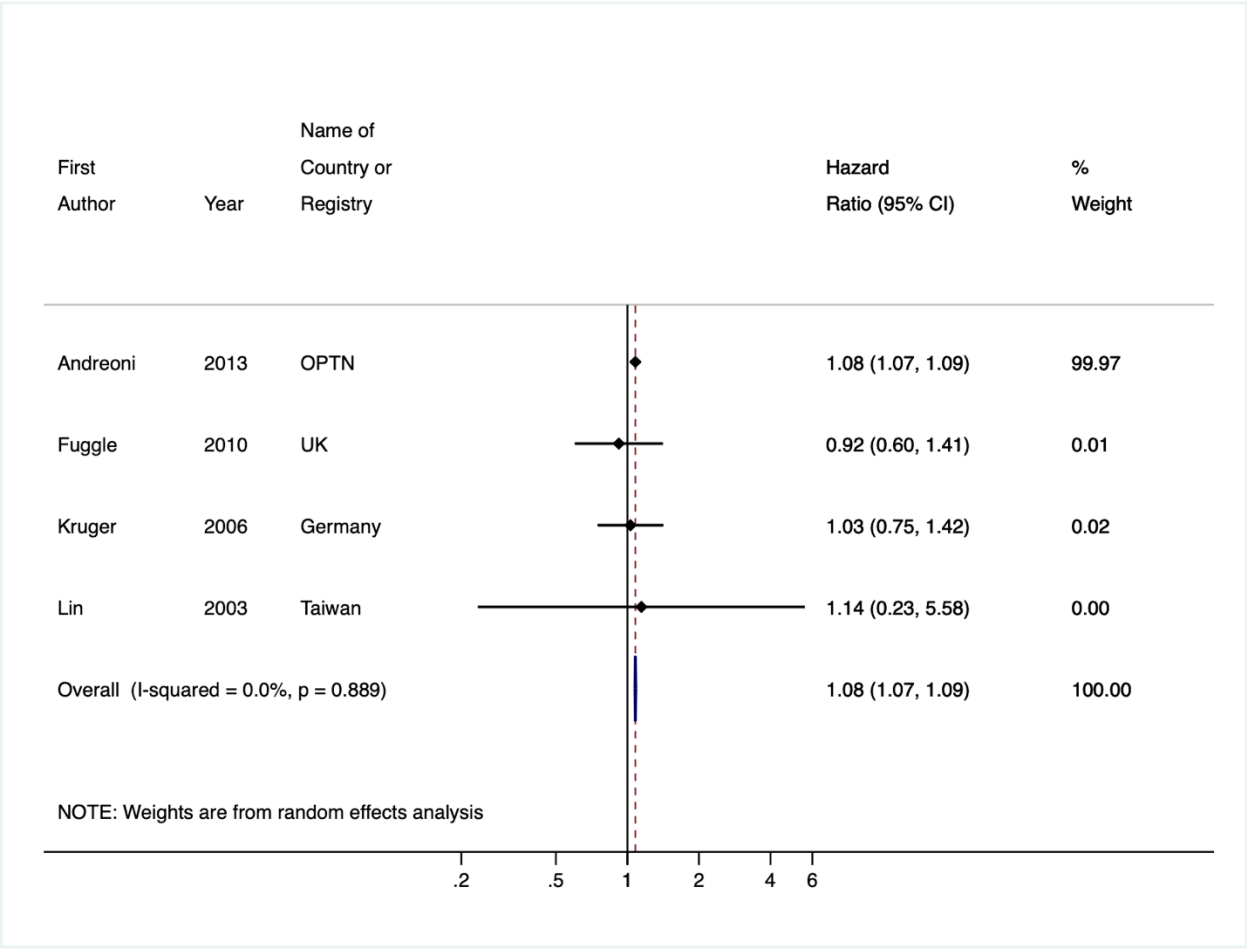
HR for every 1-unit increase

APPENDIX K – Meta-analysis of ischemic time



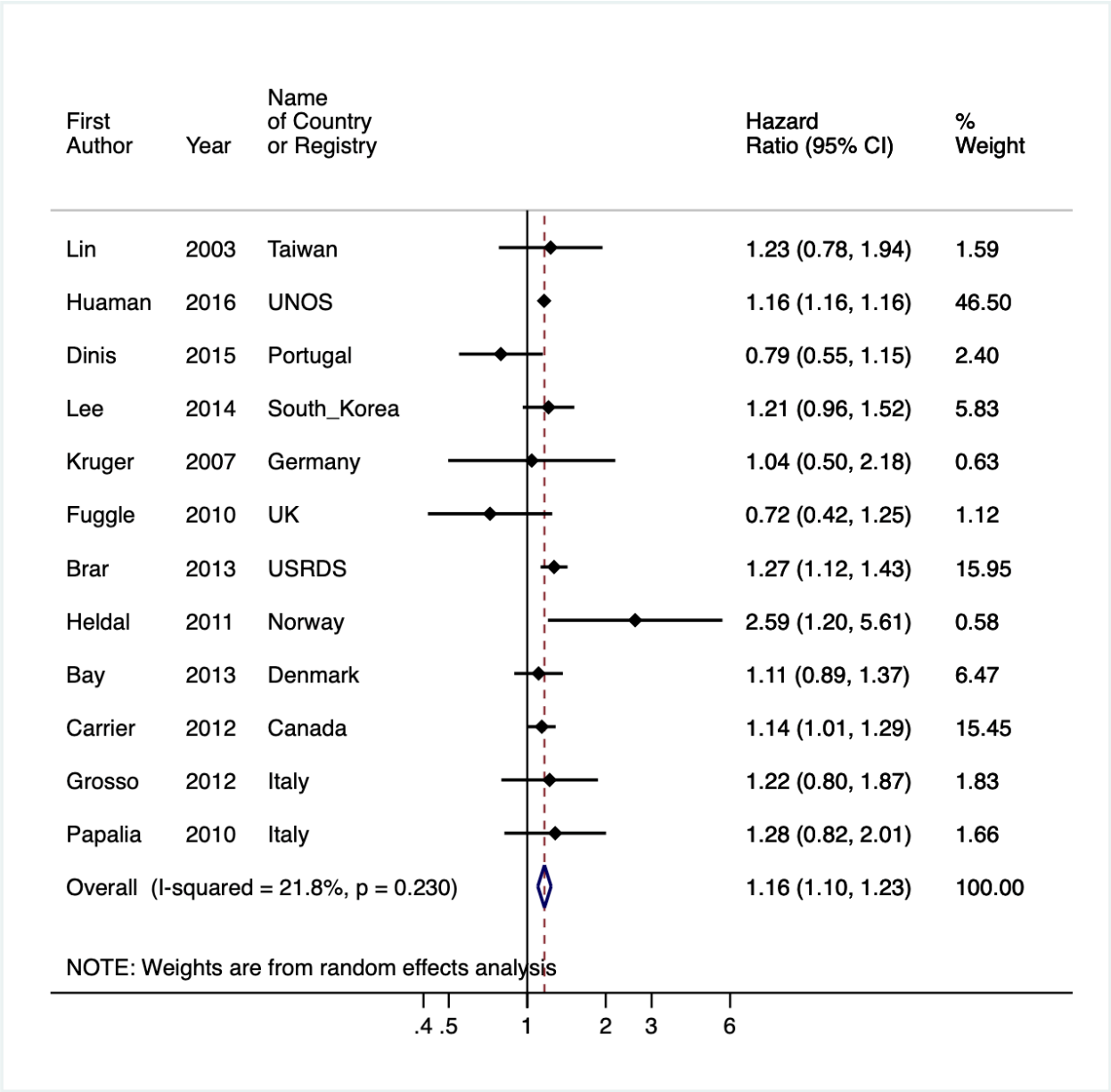
HR for every 1-hour increase in cold-ischemic time.

APPENDIX L – Meta-analysis human leukocyte antigen mismatch



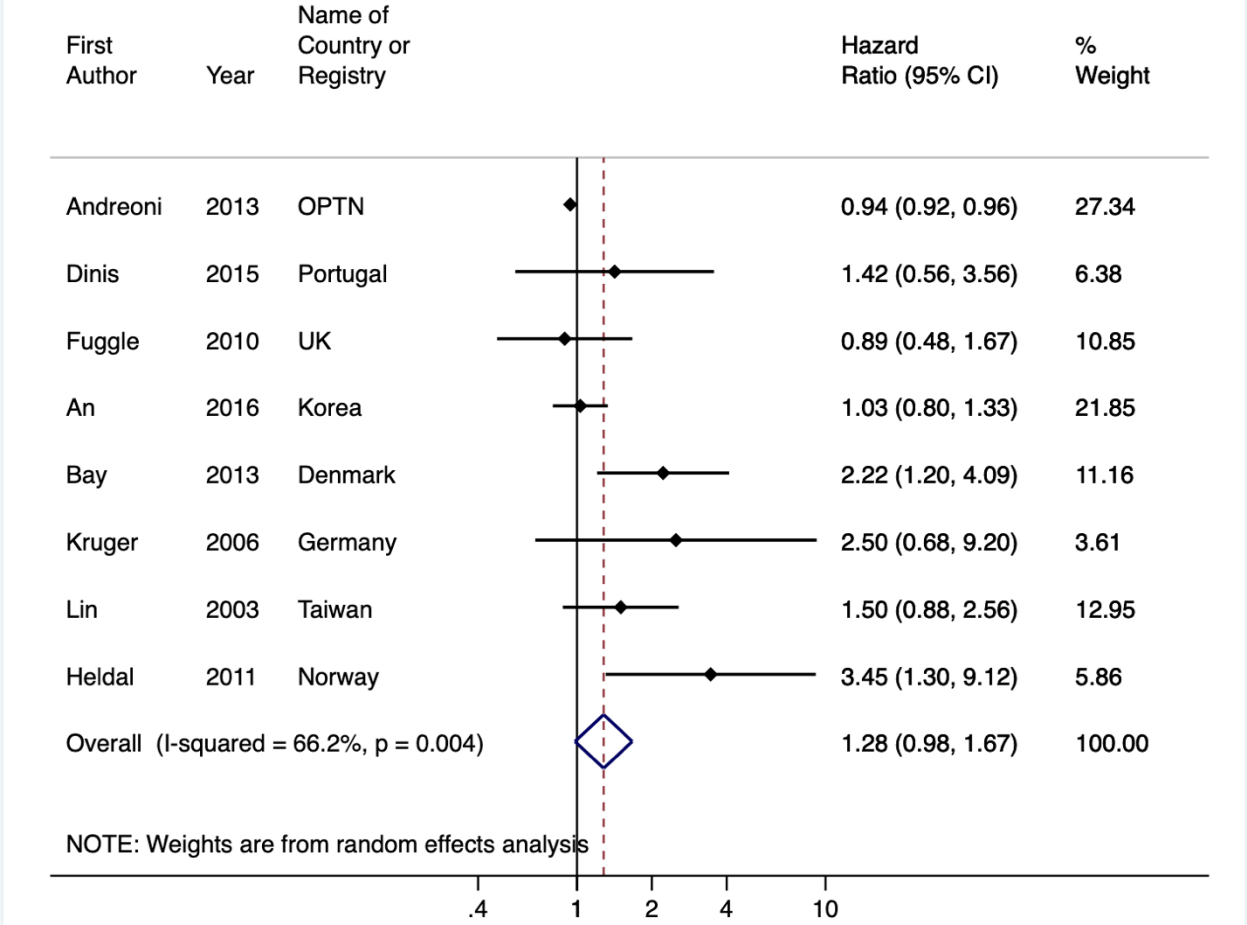
HR for ever 1 increasing HLA mismatch.

APPENDIX M – Meta-analysis recipient age



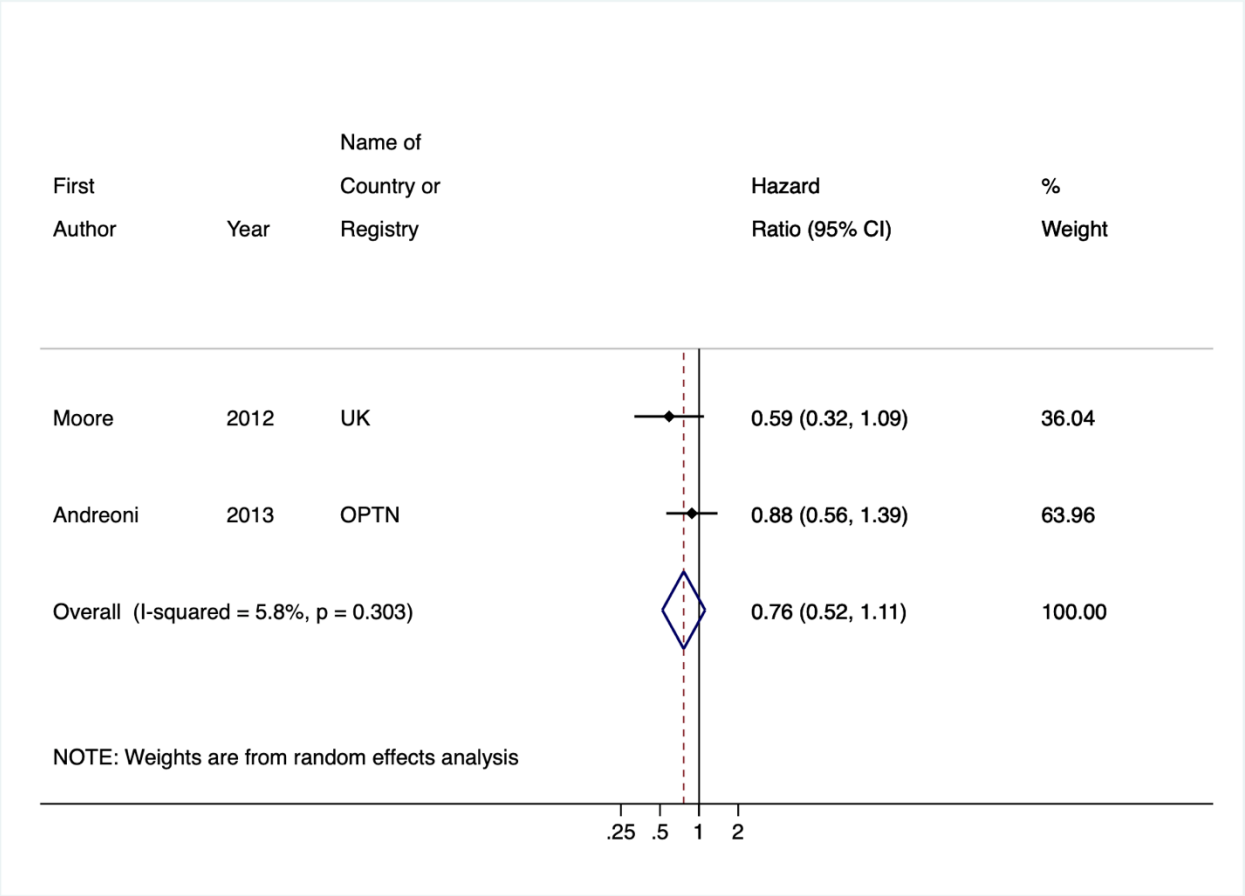
HR for every 10-year increase in recipient age

APPENDIX N – Meta-analysis of recipient sex



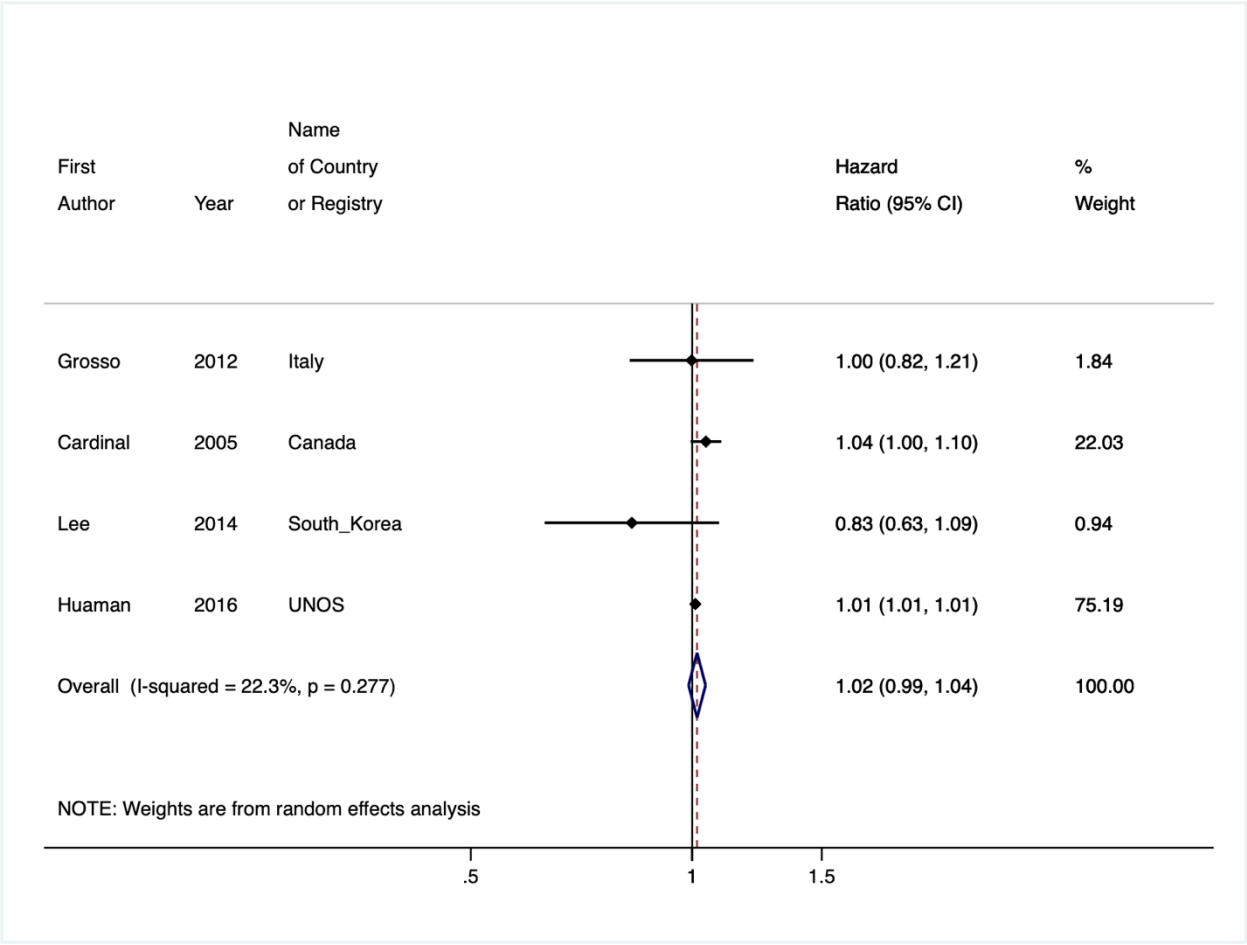
HR for male sex compared to female sex

APPENDIX O – Meta-analysis of recipient race



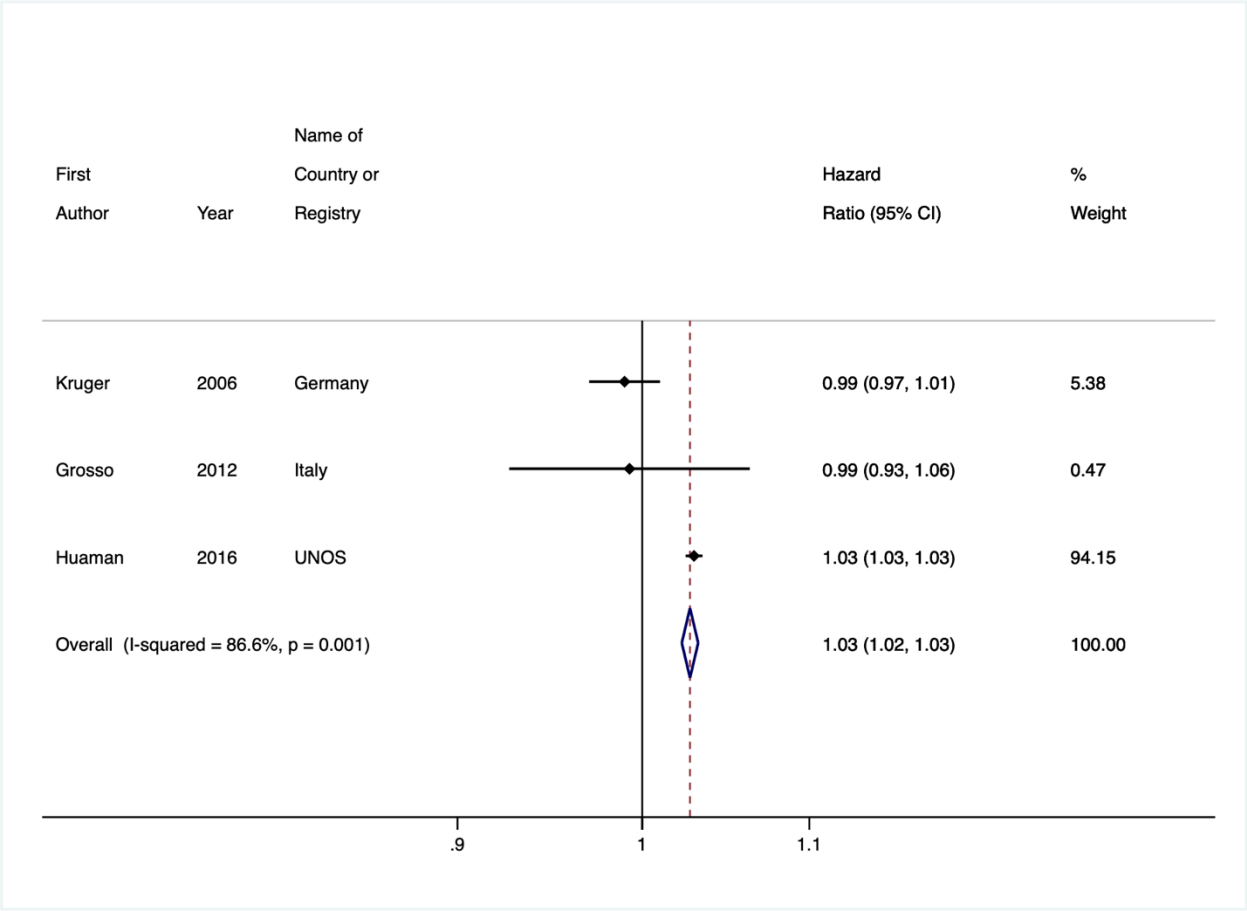
HR for Caucasian race compared to all other

APPENDIX P – Meta-analysis of recipient body mass index



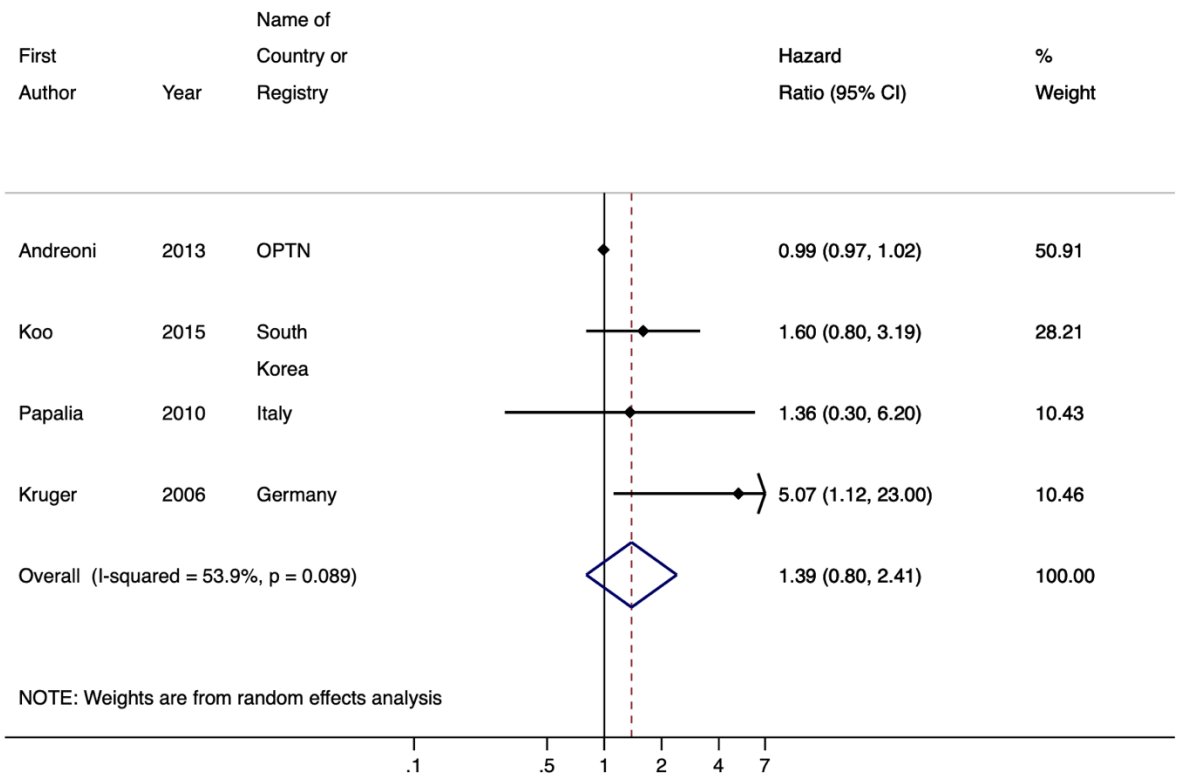
HR for every 1 kg/m^2 increase in recipient BMI

APPENDIX Q – Meta-analysis of recipient duration of time on dialysis

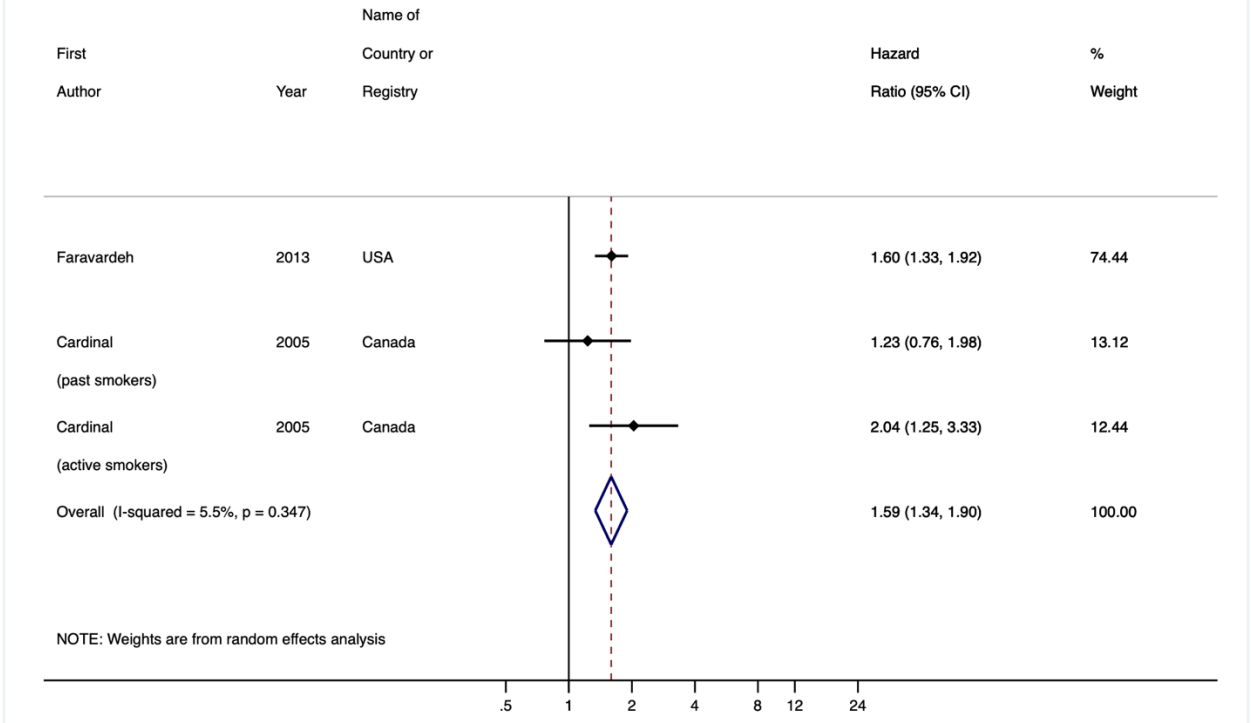


HR for every 1-year increase in duration of time on dialysis pre-transplant

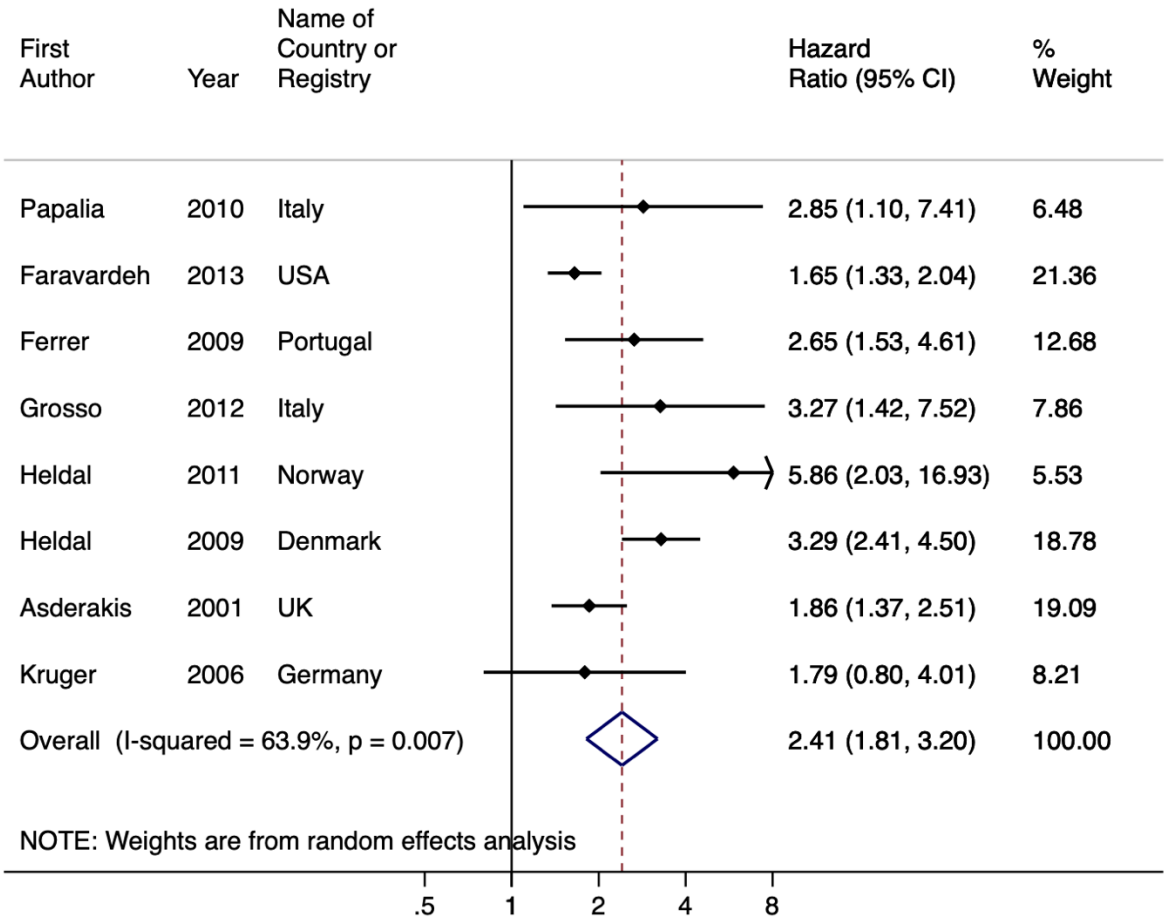
APPENDIX R – Meta-analysis of recipient diabetes



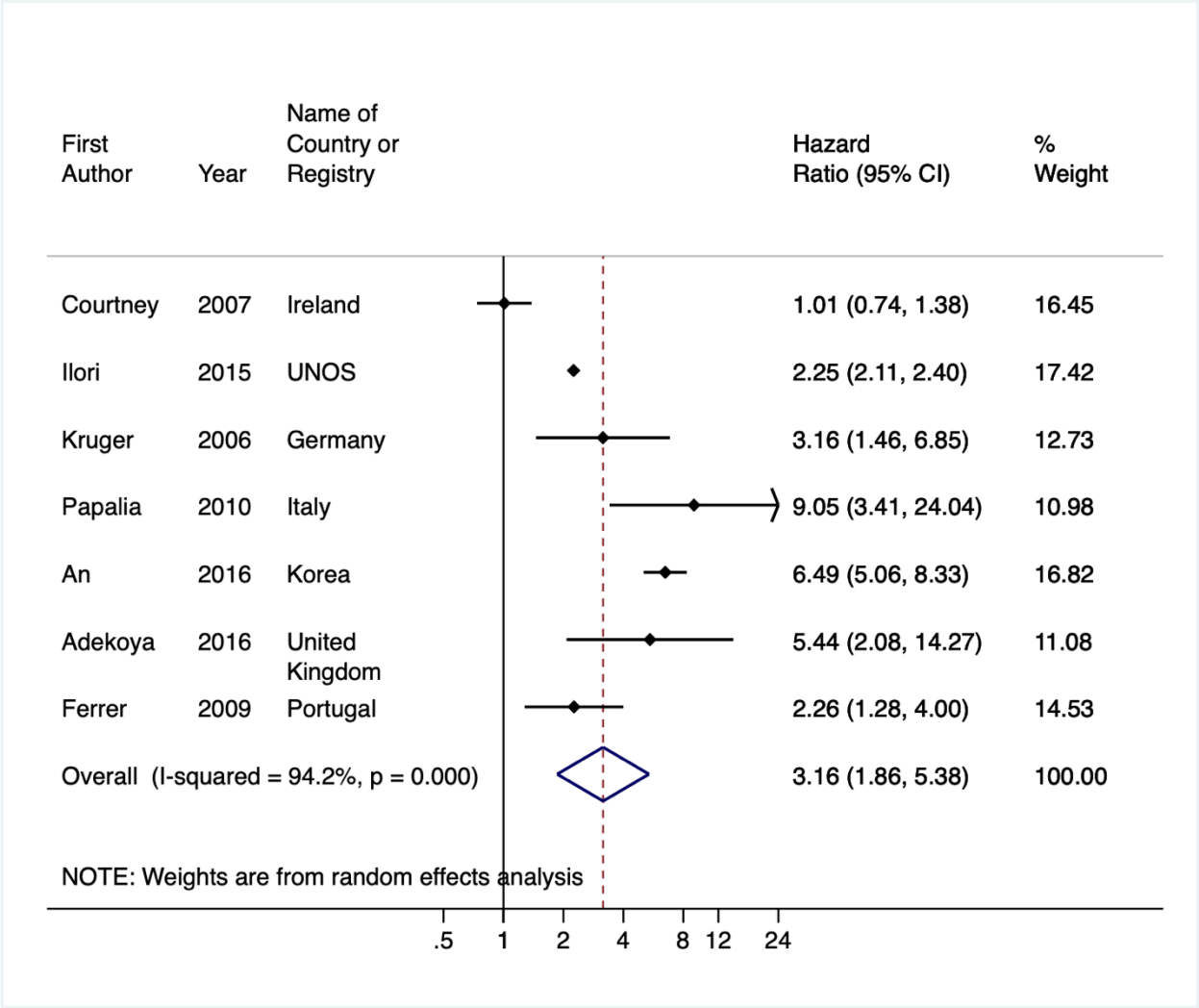
APPENDIX S – Meta-analysis of recipient history of smoking



APPENDIX T – Meta-analysis of delayed graft function

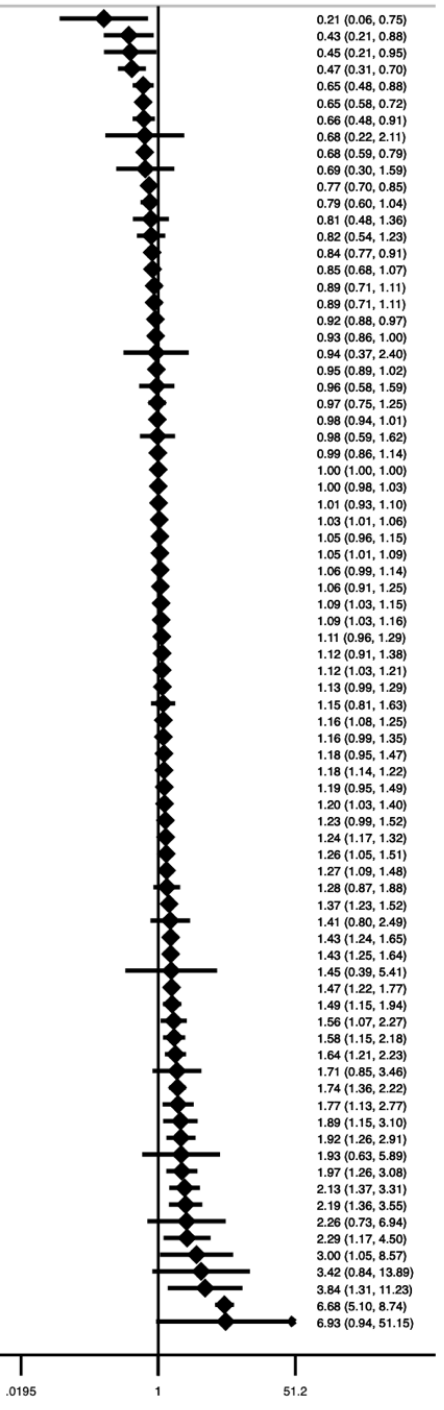


APPENDIX U – Meta-analysis of acute rejection



APPENDIX V – List of predictors identified by one individual study

First Author	Year	Name of Country or Registry	Predictor	Category	Hazard Ratio (95% CI)
Diaz	2009	Spain	Recipient Exercise	Yes vs no	0.21 (0.06, 0.75)
Bay	2013	Denmark	Recipient C4 genes	<4 vs >4	0.43 (0.21, 0.88)
Adekoya	2016	United Kingdom	Donor eGFR	>60 vs <60	0.45 (0.21, 0.95)
Molnar	2011	USA	Pre-transplant Recipient Albumin	>4.3 vs <3.77	0.47 (0.31, 0.70)
Molnar	2011	USA	Pre-transplant Recipient Albumin	3.97 - 4.13 vs <3.77	0.65 (0.48, 0.88)
Ilori	2015	UNOS	Primary renal disease	Cystic kidney disease vs diabetes	0.65 (0.58, 0.72)
Molnar	2011	USA	Pre-transplant Recipient Albumin	4.13 - 4.30 vs <3.77	0.66 (0.48, 0.91)
Grosso	2012	Italy	Donor cause of death	Traumatic vs non-traumatic	0.68 (0.22, 2.11)
Andresdottir	2005	Euro-transplant International foundation	Era transplantation	1996 - 2002 vs. 1990-1995	0.68 (0.59, 0.79)
Fuggle	2010	UK	Donor recipient relationship	Other unrelated vs Sibling	0.69 (0.30, 1.59)
Anderson	2015	OPTN	Biopsy on kidney graft	Yes vs no	0.77 (0.70, 0.85)
Molnar	2011	USA	Pre-transplant Recipient Albumin	3.77 - 3.97 vs <3.77	0.79 (0.60, 1.04)
Lin HH	2003	Taiwan	Recipient Anti-HCV Ab+	Yes vs no	0.81 (0.48, 1.36)
Dinis	2015	Portugal	Kidney weight to recipient weight	per g/Kg inrease	0.82 (0.54, 1.23)
Ilori	2015	UNOS	Primary renal disease	Glomerulonephritis vs diabetes	0.84 (0.77, 0.91)
Courtney	2007	Ireland	Donor HO-1 polymorphism	SS/SL vs LL	0.85 (0.68, 1.07)
Courtney	2007	Ireland	Donor HO-1 polymorphism	SS/SL vs LL	0.89 (0.71, 1.11)
Courtney	2007	Ireland	Renal diagnosis	NR	0.89 (0.71, 1.11)
Nee	2013	USRDS	Median household income	Per 1 level increase	0.92 (0.88, 0.97)
Ilori	2015	UNOS	Primary renal disease	Other vs diabetes	0.93 (0.86, 1.00)
Fuggle	2010	UK	Donor recipient relationship	Son/daughter vs Sibling	0.94 (0.37, 2.40)
Ilori	2015	UNOS	Primary renal disease	Hypertension vs diabetes	0.95 (0.89, 1.02)
Fuggle	2010	UK	Donor recipient relationship	Parent vs Sibling	0.96 (0.58, 1.59)
Molnar	2012	USA	Pre-transplant Alkaline Phosphatase	<80 U/L vs 80-120	0.97 (0.75, 1.25)
Nee	2013	USRDS	Education level	All	0.98 (0.94, 1.01)
Fuggle	2010	UK	Donor recipient relationship	Spouse/Partner vs Sibling	0.98 (0.59, 1.62)
Anderson	2015	OPTN	Creatinine clearance match ratio (recipient/donor)	0.70 - 0.92 vs 0.33 - 0.70	0.99 (0.86, 1.14)
Anderson	2015	OPTN	Warm ischemic time	NA	1.00 (1.00, 1.00)
Nee	2013	USRDS	Current employment status	Per 1 level increase	1.00 (0.98, 1.03)
Huaman	2016	UNOS	Donor blood culture positive	Yes vs no	1.01 (0.93, 1.10)
An	2016	Korea	Pre-transplant CV risk score	per 1 unit increase	1.03 (1.01, 1.06)
Anderson	2015	OPTN	Pumping of kidney graft	Yes vs no	1.05 (0.96, 1.15)
Andreoni	2013	OPTN	Donor hypertension	Yes vs no	1.05 (1.01, 1.09)
Ilori	2015	UNOS	Neighborhood poverty	>20% vs <=20%	1.06 (0.99, 1.14)
Nee	2013	USRDS	Education level	High school vs postsecondary	1.06 (0.91, 1.25)
Carrier	2012	Canada	Time on waiting list	Per 1 year increase	1.09 (1.03, 1.15)
Ilori	2015	UNOS	Insurance	Public and other vs Private	1.09 (1.03, 1.16)
Anderson	2015	OPTN	Creatinine clearance match ratio (recipient/donor)	1.17 - 1.55 vs 0.33 - 0.70	1.11 (0.96, 1.29)
Molnar	2012	USA	Intact parathyroid hormone	iPTH 300-500 pg/ml vs 150-300	1.12 (0.91, 1.38)
Diaz	2009	Spain	Recipient Diastolic BP	per 1 mmHg increase	1.12 (1.03, 1.21)
Anderson	2015	OPTN	Creatinine clearance match ratio (recipient/donor)	0.92 - 1.17 vs 0.33 - 0.70	1.13 (0.99, 1.29)
Molnar	2012	USA	Intact parathyroid hormone	iPTH 500-800 pg/ml vs 150-300	1.15 (0.81, 1.63)
Andreoni	2013	OPTN	Donor diabetes	Yes vs no	1.16 (1.08, 1.25)
Nee	2013	USRDS	Rejection within 12 months	Yes vs no	1.16 (0.99, 1.35)
Courtney	2007	Ireland	Renal diagnosis	NR	1.18 (0.95, 1.47)
Redfield	2016	UNOS	Donation after cardiac death	Yes vs no	1.18 (1.14, 1.22)
Courtney	2007	Ireland	Recipient HO-1 polymorphism	SS/SL vs LL	1.19 (0.95, 1.49)
Boffa	2017	United Kingdom	Donor AKI	Yes vs no	1.20 (1.03, 1.40)
Courtney	2007	Ireland	Recipient HO-1 polymorphism	SS/SL vs LL	1.23 (0.99, 1.52)
Huaman	2016	UNOS	Donor race	Black vs others	1.24 (1.17, 1.32)
Andresdottir	2005	Euro-transplant International foundation	Recipient Immunization	Immunized vs not	1.26 (1.05, 1.51)
Andresdottir	2005	Euro-transplant International foundation	Ischemic time	>24 hours vs <24 hours	1.27 (1.09, 1.48)
Molnar	2012	USA	Intact parathyroid hormone	iPTH >800 pg/ml vs 150-300	1.28 (0.87, 1.88)
Brar	2013	USRDS	Recipient peripheral vascular disease	Yes vs no	1.37 (1.23, 1.52)
Lin HH	2003	Taiwan	Recipient chronic hepatitis	Yes vs no	1.41 (0.80, 2.49)
Anderson	2015	OPTN	Creatinine clearance match ratio (recipient/donor)	1.55 - 4.32 vs 0.33 - 0.70	1.43 (1.24, 1.65)
Lynch	2009	USA	Rejection	Yes vs no	1.43 (1.25, 1.64)
Fuggle	2010	UK	Donor recipient relationship	Other related vs Sibling	1.45 (0.39, 5.41)
Anderson	2015	OPTN	Pre-transplant dialysis	Yes vs no	1.47 (1.22, 1.77)
Molnar	2012	USA	Pre-transplant Alkaline Phosphatase	120-160 U/L vs 80-120	1.49 (1.15, 1.94)
Moore	2012	UK	Donor Caveolin-1 Gene	CAV1 genotype AA vs. non AA	1.56 (1.07, 2.27)
Faravardeh	2013	USA	Recipient congenstive heart failure	Yes vs no	1.58 (1.15, 2.18)
Molnar	2012	USA	Pre-transplant Alkaline Phosphatase	>160 U/L vs 80-120	1.64 (1.21, 2.23)
Ferrer	2009	Portugal	Recipient comorbidities	Yes vs no	1.71 (0.85, 3.46)
Brar	2013	USRDS	Recipient Hypercholestrolemia	>200 mg/dL vs =<200 mg/dL	1.74 (1.36, 2.22)
Heldal	2009	Denmark	Rejection first 90 d	Yes vs No	1.77 (1.13, 2.77)
Nee	2013	USRDS	Rejection within 6 months	Yes vs no	1.89 (1.15, 3.10)
Asderakis	2001	United Kingdom	Ischemic time	>20 hours vs <20 hours	1.92 (1.26, 2.91)
Grosso	2012	Italy	Donor cause of death	Vascular vs non-vascular	1.93 (0.63, 5.89)
Moore	2012	UK	Donor Caveolin-1 Gene	CAV1 genotype AA vs. non AA	1.97 (1.26, 3.08)
An	2016	Korea	Recipient vascular disease	Yes vs no	2.13 (1.37, 3.31)
Lynch	2009	USA	Surgical site infection	Yes vs no	2.19 (1.36, 3.55)
Papalia	2010	Italy	Chronic allograft nephropathy	Yes vs no	2.26 (0.73, 6.94)
Bay	2013	Denmark	Recipient LOG2 Ficolin-3	NR	2.29 (1.17, 4.50)
Kruger	2006	Germany	CV event before Tx	Yes vs no	3.00 (1.05, 8.57)
Nanmoku	2012	Japan	Malignancy	Yes vs no	3.42 (0.84, 13.89)
Nanmoku	2012	Japan	ABO incompatibility	Yes vs no	3.84 (1.31, 11.23)
An	2016	Korea	Recurrent glomerulonephritis	Yes vs no	6.68 (5.10, 8.74)
Kruger	2006	Germany	Haplotype NOD2/CARD 15	NR	6.93 (0.94, 51.15)



Chapter 6: Use of GRADE for assessment of evidence about prognostic factors: rating certainty in identification of groups of patients with different absolute risks

STATUS: Published – Journal of Clinical Epidemiology

https://doi.org/10.1016/j.jclinepi.2019.12.023

Farid Foroutan[1,3], Gordon Guyatt[1], Victoria Zuk[1], Per Olav Vandvik[2], Ana Carolina Alba[3], Reem Mustafa[4], Robin Vernooij[5], Ingrid Rodriguez[6], Zachary Munn[7], Pavel Roshanov[8], Richard Riley[9], Stefan Schandelmaier[1], Ton Kujipers[10], Reed Siemieniuk[1], Carlos Canelo Avbar[5], Holger Schunemann[1], Alfonso Iorio[1].

(1) Department of Health Research Methods, Evidence, and Impact, McMaster University, Ontario, Canada
(2) Department of Medicine, Innlandet Hospital Trust, Division Gjøvik, Norway
(3) Ted Rogers Center for Heart Research, Toronto General Hospital, Ontario, Canada
(4) Department of Internal Medicine, The University of Kansas Medical Center, United States
(5) Iberoamerican Cochrane Centre, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain
(6) Clinical Biostatistics Unit, Hospital Universitario Ramón y Cajal, Madrid, Spain
(7) Joanna Briggs Institute, Faculty of Health and Medical Sciences, The University of Adelaide, Australia
(8) Department of Medicine, McMaster University, Ontario, Canada
(9) School of Health and Population Sciences, University of Birmingham, Birmingham, United Kingdom
(10) Department of Guideline Development, Dutch College of General Practitioners, Utrecth, Netherlands

Summary concepts

- GRADE's approach defines quality/certainty of evidence as certainty in effect estimates; this conceptualization also applies to bodies of evidence addressing overall prognosis in broadly defined populations.
- One can rate certainty of evidence not considering (non-contextualized) or considering (contextualized) the clinical context.
- Here we report how to apply GRADE to risk estimates of future events (i.e. prognosis) in groups of patients identified by a specific prognostic factor using both non-contextualized and contextualized approaches.
- For questions of prognosis, a body of observational evidence (potentially including patients enrolled in randomized controlled trials) begins as high certainty in the evidence.
- The five domains of GRADE for rating down certainty in the evidence, i.e. risk of bias, imprecision, inconsistency, indirectness and publication bias, as well as the domains for rating up, also apply to estimates of associations between prognostic factors and outcomes.
- Applying these concepts to systematic reviews of prognostic factor(s) provides a useful approach to determine the certainty of evidence regarding estimates of difference in risks captured by a prognostic factor, for both contextualised and non-contextualised situations.

ABSTRACT

The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach to rating certainty in the results of research studies was initially developed for therapeutic questions. The approach considers: study design, risk of bias; inconsistency; imprecision; indirectness, publication bias; magnitude of effect; and dose-response.

Questions about prognosis bear great relevance for decision-making in health care. Studies of prognosis can inform individuals about their likely outcome: for instance, in patients with a new diagnosis of cancer, are they likely to be alive or dead in five years. Further, prognostic studies can aid decisions in those considering treatment: for instance, is one's risk high enough to use medication to prevent cardiovascular disease. It follows that health care professionals and patients need to know how confident they can be regarding such prognostic estimates.

We have previously provided guidance for using the GRADE approach to determine certainty in estimates of future events in broad categories of patients (overall prognosis). Prognostic studies may also provide more or less robust estimates of the association between patient characteristics (such as their age, sex, and coexisting illness) and undesirable or desirable outcomes. GRADE's approach to certainty of the evidence aims to inform clinicians and patients of the trustworthiness of the estimates from systematic reviews of studies addressing such individual prognostic factors.

Introduction – Prognosis and prognostic factors

Prognosis studies address the likelihood of future health outcomes in a well-defined clinical population[1], such as individuals diagnosed with a particular cancer. Such studies often consider prognostic factors, which are particular characteristics of the patient (e.g., age, sex, biomarkers, genetic profile, or co-morbidities) that are associated with the probability of future health outcomes[2]. Different categories or values of a prognostic factor help define groups (strata) that have different average outcome risks. Understanding the average prognosis determined in groups defined by a prognostic factor may help patients to plan for their future, and allow health professionals to better manage distinct groups of individuals. In particular, it may help to balance benefits and harms associated with treatment and/or diagnostic investigations. For instance, patients with a good prognosis may be reluctant to undergo risky treatment. And, assuming a particular treatment's effect (such as a risk or odds ratio) is the same for all patients, those with a bad prognosis can expect a greater absolute benefit and may therefore be more willing to incur risks associated with treatment. In making such judgments, patients will be appropriately interested in the trustworthiness of the prognostic information they receive.

In a previous paper we discussed three categories of prognostic studies (panel 1), and focused on rating the certainty of evidence for studies of prognosis of a typical patient from a broadly defined population[3]; so-called overall prognosis. Here we discuss the certainty of evidence regarding prognosis in groups of patients defined by prognostic factors. In the future, we will address the certainty of evidence in studies of prognostic models and risk prediction tools.

Most often, patients present a constellation of prognostic factors and, accordingly, clinicians or health care providers rarely use single prognostic factors in clinical decision making. Nevertheless, we have focused this GRADE guidance on a single prognostic factor for one main reason: within a single study and between studies, the certainty of evidence is likely to differ across prognostic factors. For instance, within a study the confidence intervals may be wide for one prognostic factor and narrow for another. Across studies, results may be consistent for one prognostic variable and show large inconsistency for another. Thus, each prognostic factor requires a separate certainty assessment. Additionally, there are clinical situations in which a single patient characteristic (e.g. age or ethnicity) or a single disease marker (e.g. the stage of cancer) dominate clinical decision making and patient counselling.

One can think of two broad categories of use of single prognostic factors. The first is in relation to study planning and analysis: stratification of randomization, adjusted analysis, and developing a prognostic model. Following previous GRADE guidance, we refer to these situations as non-contextualized[4], which implies that there is no direct clinical action associated with the prognostic factor evidence. The second is in relation to clinical decision-making that we refer to as contextualized, which implies that clinicians and patients need risk estimates defined by the prognostic factor to inform their decisions.

We offer the guidance in this article both to those conducting systematic reviews of prognostic factors and those interested, from a clinical perspective, in the certainty of evidence supporting prognostic factor estimates from a body of evidence. The latter group will include guideline

developers – for instance, a guideline panel's confidence in a strong association between a prognostic factor and an outcome may influence their decision on whether it is appropriate to issue a specific recommendation for a subset of patients. For the non-contextualized ratings, interested parties will include statisticians and investigators considering using prognostic factors in the design and analysis of healthcare research, or in the subsequent development of a prognostic model containing multiple prognostic factors. As our discussion will reveal, non-contextualized and contextualized ratings are very unlikely to differ for risk of bias or publication bias, may differ for ratings of directness or consistency, and are most likely to differ for ratings of precision. Those using this article to evaluate an existing systematic review may find that authors did not provide all the necessary information, and may therefore need to consult the primary studies included in the review.

Methods of Development of this Guidance

We identified potential systematic reviews by searching PubMed, using the Clinical Queries filters to select SR about prognosis and prognostic factors.

We searched for and included systematic reviews that met all of the following criteria:

a) Summarizing evidence from studies addressing binary outcomes (the prognostic factors, however, could be ordinal [e.g. severity of disease] or continuous [e.g. age]).

b) Summarizing evidence on any individual prognostic factor(s) (including but not limited to: age, time, sex, race, body mass index, comorbidities, and diagnostic test results,).

c) Summarizing evidence either using aggregate data or individual patient data.

d) Report a risk of bias assessment of the included studies.

To allow for the possibility of making our judgments for both the contextualized and non-contextualized framework we excluded systematic reviews not reporting at least one absolute measure of risk, either for those with or without the prognostic factor (for binary prognostic factor) under study, or the average risk in the population (For continuous prognostic factors).

For each of the 10 chosen reviews, 14 members of the GRADE prognosis working group independently completed evidence profiles for a designated prognostic factor from each review (Appendix 1). Participants reviewed the agreement and disagreement of assessments during 90-minute teleconference meetings, and came to consensus regarding judgments and the implications of the judgments for GRADE guidance.

Key Considerations in Addressing Certainty of Evidence for Individual Prognostic Factors

As with all GRADE guidance, in examining the prognostic effect of individual factors we are addressing evidence ideally summarized in the form of systematic reviews. Fortunately, as for overall prognosis, systematic reviews addressing prognostic effect and value of individual factors are increasingly available[5-7]. In addition, investigators interested in conducting systematic reviews of prognostic factors now have access to guides for their conduct[8].

Prognostic factors are most often assessed with observational studies. Randomized trials become necessary when interventions that modify the prognostic factor become available (e.g. antihypertensive treatment for hypertension; statins for hypercholesterolemia; weight loss for obesity). When we consider the usefulness of prognostic factors in identifying those with more and less desirable outcomes, determining whether the associations are or are not causal is irrelevant. To this end randomized trials are not the optimal design.

Applying GRADE principles to questions about prognostic factors

A series of articles published in the Journal of Clinical Epidemiology provides details of the GRADE approach for rating certainty of evidence regarding treatments[9] and diagnostic tests[10], and our previous paper provides guidance regarding overall prognosis of broad populations[3]. In each case, the GRADE approach involves consideration of eight domains that may affect the certainty in the evidence; decrease in certainty: risk of bias, inconsistency, indirectness, imprecision, publication bias; increase in certainty: large effect, dose response, and plausible confounding. Depending on the study design and issues relating to these domains, certainty is ultimately designated as high, moderate, low, or very low.

The present paper provides specific guidance regarding the five domains proposed by GRADE, where limitations may decrease certainty in the body of evidence for the association between prognostic risk factors and patient important outcomes, and three domains that increase certainty. We will use examples from the systematic reviews we assessed to illustrate principles in application of GRADE to bodies of evidence on a single prognostic factor (Table 1) and provide an example of summary of findings tables including evidence profile for both non-contextualized and contextualized ratings (supplemental material). Table 2 presents the GRADE interpretation of its four levels of evidence applied to prognostic factor studies.

Risk of bias

The ideal study design

Best evidence regarding prognostic factors usually originates from observational studies (cohort studies, registries, or database linkage studies) and, for both studies of overall prognosis, and

studies of individual prognostic factors, such studies start with high certainty ratings[3]. Although secondary analyses of randomized control trials (RCTs) can also provide evidence regarding prognosis, observational studies typically yield higher certainty because eligibility criteria for RCTs usually include filters (e.g. restrictions in age, comorbidity / performance status, drug intolerance) that exclude patients likely to be relevant for assessment of prognostic factors. Moreover, eligible patients may decline to participate in RCTs, and their reasons for declining may be related to their prognosis. For example, a study that compared differences in characteristics of 4713 enrolled patients in the Euro Heart Survey on Coronary Revascularization to 8647 patients enrolled in 14 major RCTs comparing percutaneous coronary intervention with coronary artery bypass grafting[11] reported that patients enrolled in trials were significantly younger, and less frequently suffered from hypertension, hyperlipidemia, diabetes, peripheral vascular disease, and cerebrovascular disease. If investigators do use RCT data to address prognosis, they may use either the control group alone, or the entire RCT cohort, in which case adjustment for the intervention will be required.

Assessing risk of bias in individual studies

When evaluating the risk of bias regarding the association between a prognostic factor and an outcome, we are concerned about elements in study design and conduct that may result in over- or underestimation of the true effect on prognosis as measured by a risk ratio, hazard ratio or odds ratio. The QUality In Prognosis Studies (QUIPS) provides a useful instrument developed for evaluation of risk of bias in studies addressing prognostic factors[12]. Risk of bias instruments such as Prediction model Risk Of Bias ASsessment Tool (PROBAST) may also provide useful

considerations for assessment of risk of bias[13]. Chapter 13 of the Cochrane Handbook for Systematic Reviews of Interventions provides additional indirect guidance[14].

Adjusted and unadjusted associations

When evaluating individual prognostic factors, we are interested in their association with patient important outcomes. In most instances, prognostic factors are correlated with each other, and as a result their individual associations with outcomes may be potentially misleading.

For instance, consider a critical care physician estimating a patient's risk of serious bleeding as a result of gastric stress ulceration. The clinician will find that – when considered individually – mechanical ventilation, coagulopathy, hepatic failure, sepsis and hypotension all increase the risk of bleeding by fivefold or more[15]. The clinician facing a patient in whom all these factors exist would be tempted to conclude that the risk of bleeding is extremely high.

A multivariable adjusted analysis, however, revealed that only the first two of these are independent prognostic factors – the others derive their apparent predictive power from their association with the first two, and clinicians should consider only mechanical ventilation and coagulopathy in assessing risk of bleeding. Thus, when clinicians consider multiple factors simultaneously in making prognostic estimates, adjustment in multivariable analyses including all factors is required to generate unbiased and useful estimates of added prognostic value. The same is true for clinical investigators using prognostic studies for stratification, adjustment, or prediction guide construction. This is sometimes referred to as the 'independent' or 'adjusted' prognostic value of a prognostic factor[2]. Aside from standard multiple regression, 'independent'

or ‘adjusted’ prognostic value may also be derived from score matching, inverse probability weighting, marginal structural modeling, and machine learning.

The situation differs, however, in a less common scenario in which clinicians rely on a single easily measured prognostic factor that is related to numerous less important prognostic factors, and is as good – or almost as good - as an overall model including all factors. Consider, for example clinicians assessing patients for the need of diagnostic imaging to rule out venous thromboembolism (VTE). Here individuals with a negative D-dimer alone have an expected 98.9% probability of uneventful follow up over three months [16 17], almost identical to the 98.6% likelihood of not having a clot for patients with a Wells clinical predictive model score (considering 7 prognostic factors) of ≤ 1[18]. In this case, in patients with a negative D-dimer, the unadjusted estimate provides essentially the same level of information as the model that includes all relevant variables.

This latter situation is, however, unusual; therefore, in the remainder of this paper, we will consider the more common context in which, in their contextualized setting, clinicians simultaneously consider the impact of multiple factors on patients’ prognosis, and so should examine the adjusted effect of a prognostic factor. In such situations lack of adjustment for existing (or established) prognostic factors represents a source of bias.

For instance, Sanchez et al. conducted a systematic review evaluating the prognostic association between right ventricular (RV) dysfunction and in-hospital mortality in patients with acute

pulmonary embolism[19]. The authors included seven studies with 666 patients, observed between 2000 to 2006, in their meta-analysis, and reported a relative risk of 2.43 (95% CI of 1.33 to 4.45), suggesting that the risk is 2.43 times larger in those with RV dysfunction compared to without.

The authors inform us that only one identified study evaluated the adjusted association between RV dysfunction and mortality, but do not make it explicit whether they are reporting the adjusted or unadjusted estimate within the meta-analysis. Within the forest plot, however, for all seven studies the authors report the number of patients with the outcome in the RV dysfunction and non-RV dysfunction group, hence working backwards, we can estimate each individual study's relative risk included in the meta-analysis. As our estimation matches that of the forest plot we can, therefore, infer that all relative risks being pooled are unadjusted. Clinicians, in addressing the prognosis of patients with pulmonary embolus, will consider factors other than RV dysfunction, including systolic blood pressure, the extent of hypoxemia, and heart and respiratory rates. Therefore, adjusted analysis is crucial to avoid a misleading conclusion for this assessment. Indeed, in contrast to the unadjusted estimate suggesting association beyond chance, the authors of the only study that did adjust for other predictors failed to observe a statistically significant association between RV dysfunction and mortality. Thus, for both non-contextualized and contextualized ratings, the study's failure to adjust requires rating down for risk of bias.

The discussion thus far has made evident that studies of prognostic factors should, ideally, conduct a multivariable analysis that includes all prognostic factors associated with the outcome

of interest. In most instances, however, the number and choice of factor adjustment will vary across studies. If the literature is dominated by studies that fail to adjust for one or more crucial predictors, adjusted estimates may still be at high risk of bias.

Results may influence risk of bias judgements

Another consideration reflects GRADE guidance for intervention questions. Prior to rating down for risk of bias, authors of systematic reviews should determine if studies at high risk of bias have actually biased the meta-analysis results. If a body of evidence includes a robust collection of studies at both high and low risk of bias, and a subgroup analysis shows that the high risk of bias studies differ importantly in their estimates from the low risk of bias studies, one should rely only on estimates from the latter. If, however, studies at high and low risk of bias provide similar estimates of association, authors can narrow confidence intervals by including all studies and not rate down for risk of bias. In making this judgement, authors need to consider the weight that each study contributes to the final estimate of effect when considering the impact of including studies at higher risk of bias.

For instance, the review by Cheng et al. included studies evaluating the association between smoking and onset of venous thromboembolism (VTE) in patient at risk for VTE[20], and reported a relative risk (RR) of 1.19 (95% CI of 1.15 to 1.22). The authors report that some studies adjusted only for body mass index, others for body mass index, cholesterol, diabetes, hypertension, alcohol consumption, and physical activity. This observed variation raises the possibility that a more comprehensive adjustment would provide less biased results. Visual inspection of the point

estimates and 95% confidence interval, however, shows minimal difference between studies with more or less comprehensive adjustment. In this instance, the similar associations suggest that the extent of adjustment is unimportant, and there is no need to rate down for risk of bias.

In the same review, the authors included studies that utilized different criteria for diagnosis of VTE[20]. Two of the largest studies following over 3 million individuals utilized physician billing codes for determining cases of VTE, whereas the smaller studies utilized radiologic criteria for VTE diagnosis. The administrative physician billing codes are likely to be far less accurate than explicit radiologic criteria and thus represent a potentially important source of bias. In this systematic review, however, the two high risk of bias studies again provide similar estimates for the association between smoking and VTE as the low risk of bias studies. One can therefore include all studies without need to rate down for risk of bias.

Inconsistency

GRADE considerations for judging inconsistency (heterogeneity) in prognostic and therapeutic studies are similar and include variability in point estimates, extent of overlap of confidence intervals, and – for contextualized ratings - where absolute risk point estimates lie in relation to clinical decision thresholds[4]. There can, however, be one important difference; statistical measures of heterogeneity, such as the I^2, are much less useful when large studies are involved, which may be the case for observational studies addressing prognosis. Here confidence intervals are frequently narrow which may result in high I^2, implying statistical heterogeneity in the absence of what would constitute important inconsistency.[3]

Between-study heterogeneity may be more prevalent in prognostic factor research, and might be less concerning than in reviews of treatment effects. In prognostic factor studies there is much more variation in designs (e.g. cohort study data, randomised trial data, routine care registry data and case control study data may all be used in one meta-analysis), patient inclusion criteria, prognostic factor and outcome measurement, follow up time, methods of statistical analysis, and in the adjustment of (and number of) other prognostic factors[8].

Systematic review authors should nevertheless prepare for substantial inconsistency by generating a priori hypotheses that may explain the heterogeneity they encounter. Reviewers may define substantial heterogeneity through visual inspection of individual point estimate and 95% CI of individual studies in relation to decision thresholds. In the non-contextualized setting, if reviewers are rating the certainty that the prognostic effect (i.e. risk ratio, odds ratio or hazard ratio) varies from 1.0, they will not rate down as long as all studies suggest some degree of association, whatever the magnitude. In the contextualized setting, authors may conclude substantial heterogeneity when the course of action taken through the consideration of prognosis is liable to differ between the individual studies.

When substantial heterogeneity is observed, authors should determine the extent to which their hypotheses explain the inconsistent results. Only when such exploration proves fruitless and substantial unexplained heterogeneity remains, should authors consider rating down the certainty of the evidence for inconsistency.

For instance, Witlox et al. investigated the association between delirium and post-discharge mortality in a meta-analysis of 7 observational studies each of which reported an adjusted hazard ratio[21] and reported that the presence of delirium was associated with a close to doubling of the hazard of dying (HR 1.95; 95% CI 1.51 to 2.52). In six of the eligible studies, the 95% confidence intervals varied widely. One study, however, reported a notably higher hazard ratio, with minimal overlap of 95% confidence interval with the other studies (HR 4.04; 95% CI 2.19 to 7.46).

In deciding whether to rate down in such a situation, one should consider the contribution of the aberrant study to the pooled estimate – in this case responsible for only 12% of the weight. The low weight suggests that rating down for inconsistency is unnecessary, a conclusion supported by a sensitivity analysis demonstrating similar pooled estimates whether or not one includes the aberrant study (figure 1).

Responding to inconsistency

If associations differ between subgroups of patients identified by a characteristic or prognostic factor other than that under consideration, evidence that includes within-study comparisons, a very low p-value for the test of interaction ($\leq$0.01), a small number of a priori hypotheses ($\leq$3) with a correctly specified direction, and compelling indirect evidence (biological rationale), supports an inference that the subgroup effect is real rather than spurious[22]. In such a situation, the authors should report the separate estimates of the association between the prognostic factor and outcome for the relevant subgroups, thus diminishing or resolving the inconsistency.

Imprecision

Judging imprecision represents the key area in which non-contextualized and contextualized ratings are liable to differ. In the former setting, authors will rate their certainty that the prognostic effect as measured by a risk (odds or hazard) ratio differs from 1.0. They may choose a different threshold – such as a relative effect of 1.5, but finding a rationale for such an alternative threshold is likely to be challenging, and we will restrict our discussion to a threshold of 1.0.

In the contextualized setting, authors will make their decision regarding precision based on the relation between the confidence interval and a clinical decision threshold. If the clinical action would not change when the estimate is at the lower versus higher boundary of the confidence interval, the risk estimate is sufficiently precise, irrespective of the width of the interval.

Because the implications of relative estimates will differ depending on baseline risk (a RR of 2.0 may increase one's risk from 1 to 2% or 20% to 40%) clinical decision thresholds must be expressed as absolute measures of risk. As a consequence, the most efficient and directly applicable measure of association for a prognostic factor would be an absolute risk difference. If the boundaries of the confidence interval around the absolute risk difference lie on the same side of the decisional threshold, there is no important imprecision and no reason for downgrading the certainty of the evidence.

This guidance raises a challenge: prognostic factor studies typically focus on relative rather than absolute measures of association between the prognostic factor and the outcome, and often do

not report absolute measures. When this is the case, applying GRADE guidance requires converting the relative measure into an absolute one, which in turn requires an estimate of the absolute risk of event in those without the prognostic factor.

In the setting of prognostic factor research, baseline risk is the risk of outcome in the subset of patients without the prognostic factor of interest, or, when multiple prognostic factors are simultaneously addressed, in the subset of patients with the lowest risk. Often studies do not report an estimate of this baseline risk, nor the number of events and persons at risk. Rather, authors typically report the total number of events observed in the overall population irrespective of any prognostic factor.

If our interest is in the typical prognosis in an overall population, such a report would provide the information required – the average risk. Deriving the baseline risk in patients without a specific risk factor from this average risk requires considering both the prevalence of the prognostic factor and the strength of the association. Box 1 provides a worked example and a simple formula for the calculation. A freely available calculator for calculation of the absolute risks can be found at the following address: http://hiru.mcmaster.ca/AbsoluteRiskCalculator/. In a separate paper, we provide a more detailed discussion of the topic, including formulas applicable to binary, ordinal and continuous variables[23].

Box 1 – how to calculate absolute risk from prevalence and relative risk

Wiltox et al. studied the association between delirium and mortality[21]. From the supplementary material of the systematic review, we can calculate a prevalence for 31% for delirium. The average risk of mortality (i.e. the risk in the entire population, irrespective of the patients having or not the prognostic factor) amongst the 7 studies is 185 per 1000 persons. The relative risk of mortality in persons with delirium as compared to those without is 1.55 (95% CI 1.17 to 1.89). The absolute risk in patients with delirium will be equal to $\frac{245}{1000}$. The risk in those without delirium will be $\frac{158}{1000}$.

If no measure of absolute risk is available in the systematic review, one would need to rely on external evidence, for example from a body of studies of overall prognosis, providing either a baseline or an average risk.

Consider a systematic review and meta-analysis of predictors for graft loss (defined as re-transplantation or return to dialysis) after kidney transplantation, authors identified delayed graft function as a significant prognostic factor[24]. The pooled hazard ratio from 4 observational studies suggested an 89% increase in risk of graft loss in patients with delayed graft function (HR 1.89, 95% CI of 1.46 to 2.47). In the non-contextualized setting, the confidence interval excludes a HR of 1.0 by a considerable margin, indicating there is no need to rate down for imprecision.

The contextualized setting may involve additional considerations. The lower the risk of the outcome in the entire cohort, and the more prevalent the prognostic factor, the lower the likelihood that the prognostic factor, even strongly associated with the outcome, will translate into an important difference in risk. The United Nations of Organ Sharing (UNOS) provides an average 1-year graft loss risk of 74 cases per 1000 patients. Among the 4 meta-analyzed studies, the review authors observed a prevalence of 35% for delayed graft function.

Using these values and the calculator referenced above, the absolute risk of 1-year graft loss in patients without delayed graft function is 60 per 1000. In those with delayed graft function, the absolute risk increases to 100 per 1000 (40 more, 95% CI of 30 more to 70 more). The contextualized setting requires asking: might clinical decisions differ between a 3% and a 7% likelihood of graft loss. Perhaps not, and if so there is no need to rate down for imprecision.

Contrast this with results related to a much rarer prognostic factor, strongly associated with graft loss. The same systematic review identified acute rejection to increase the risk of graft loss by 3-fold (HR 3.16, 95% CI 1.86 to 5.38). Here again, the decision not to rate down for imprecision in the non-contextualized setting is clear.

Within the contextualized setting, however, the prevalence of acute rejection is only 7%. The absolute risk of 1-year graft loss in patients without acute rejection is 70 per 1000 patients. In those with acute rejection, the absolute risk increases to 190 per 1000 (120 more, 95% CI of 60

more to 210 more). Might one make different decisions at a likelihood of graft loss of 6.0% versus 21%? Perhaps so, and if that is the case one would rate down for imprecision.

The authors of the systematic review chose 5% (50 per 1000) as the risk difference threshold that might influence clinical decision-making. If one accepted that threshold, one would rate down for imprecision when considering delayed graft function, but not rate down when considering acute rejection.

The final decision regarding appropriate threshold of risk for decision-making may differ for the systematic review authors and the guideline panels resulting in varying judgments for imprecision. If systematic reviewers present all the data needed to assess the evidence they report against a threshold it will then be possible – indeed, potentially very useful - to examine the implications of different plausible thresholds. Reporting absolute risk in people with or without the prognostic factors represents the most critical step.

Indirectness

Systematic review authors, guideline developers, and other evidence users need to consider whether the populations and outcomes studied correspond to their population and outcome of interest. GRADE refers to these issues, sometimes labelled as generalizability or applicability, as issues of directness.

For prognostic factors, indirectness might originate when the care provided in a target population is sufficiently different from the way the patients were managed in studies included in the SR one

is appraising. In the non-contextualized setting, the target population might be those entered in a clinical trial, those included in an observational study addressing intervention impact, or the potential target population for a clinical decision guide. In the contextualized setting, one would focus on the clinical target group. Considerations are similar if the studied outcome is not fully representative of the outcome of interest, the uncertainty may represent a problem with indirectness.

In the systematic review addressing prognostic factors for graft loss at 1-year post kidney transplantation, authors meta-analyzed two cohort studies examining the association between donor's creatinine and 1-year graft loss. The loss of grafted kidney does not – fortunately – mean the patient is more likely to die. The best available evidence for the association between donor creatinine and graft loss, however, did not separate mortality from graft loss. Authors of the individual studies combined return to dialysis and re-transplantation with all-cause mortality. The systematic review authors observed a pooled hazard ratio of 0.95 (95% CI 0.84 to 1.07) for every 1 mg/dL increase in serum creatinine. The inclusion of mortality in the composite outcome may have contributed to the failure to demonstrate an association, a concern that warrants rating down for indirectness both for a clinical trialist designing an RCT studying an intervention to prevent graft loss, and the clinician counselling a patient regarding the likelihood of graft loss.

Publication bias

There is evidence suggesting that publication bias is as frequently a problem for prognostic factor research as for any other research field[2 25-28]. One possible approach to obtain evidence at lower risk of overestimating associations would be to search for studies in which the prognostic factor

of interest is one among many other factors assessed – that is, studies focusing on a number of predictors, as opposed to one single predictor alone. Selective reporting, however, may still bias evidence from studies reporting multiple prognostic factors[2].

As is the case for overall prognosis of large groups, tests for small study effects – of which publication bias is one cause - that normalize the distribution (e.g. Begg's test[29]) may be most useful. Other tests, e.g. Debray's test[30], Peter's tests[31], and Egger's test[32], are applicable when heterogeneity is low and data are normally distributed.

Beyond using a statistical test, careful visual exploration of the funnel plot may be helpful. For example, a systematic review by Vasilevska et al. investigated the risk of cervical cancer in indigenous versus non-indigenous women[33] and found a relative risk of 2.11 (95% CI 1.60 to 2.78). The authors concluded publication bias based on a positive Egger's test for the outcome "cervical cancer not otherwise specified". The funnel plot demonstrates, however, that the missing small studies are those that, had they been present, would show strong positive associations (figure 2). If one believed that selective publication of small negative studies is unlikely (an assumption that strikes us as reasonable, given that neither authors with small studies failing to find an association, nor editors considering a manuscript with that conclusion, are likely to be enthusiastic regarding publication) , one would disregard the positive Egger's test and not rate down for publication bias. Such considerations apply to both non-contextualized and contextualized settings.

Rating up certainty

GRADE's criteria for rating up certainty in treatment studies includes large effect, a dose response gradient, or situations where all plausible confounders or biases would decrease an apparent treatment effect or create a spurious effect when results suggest no effect. Analogous situations might be envisioned for systematic reviews addressing prognosis factors.

Thus far, we found no examples of systematic reviews in which we can rate up for prognostic factors based on dose response, large effect or for the nature of plausible biases. Examples may emerge with further use of GRADE criteria for prognostic studies of specific prognostic factors. Those interested may refer to GRADE guidance for therapeutic questions for further discussion on these domains and examples in which one may be warranted to increase certainty. One may imagine a non-contextualized scenario in which we observe a very strong association (>5 or <0.2) with no concerns of risk of bias or imprecision. Under such circumstances, it may be warranted to increase certainty in the evidence. In our experience, however, we did not observe any example of such nature.

Additional/Cautionary remarks

Some special considerations for application of GRADE domains arise when simultaneously considering more than one GRADE domain. These considerations apply equally to non-contextualized and contextualized ratings.

For instance, risk of bias and inconsistency may be correlated with one another. One potential hypothesis for exploration of inconsistency can be risk of bias. It may be that authors observe no

significant difference in the effect estimate observed from studies classified as high and low risk of bias. Therefore, authors may opt to maintain high risk of bias studies within their analysis. With regards to inconsistency, however, authors may decrease their certainty in the evidence due to high inconsistency amongst high risk of bias studies. In such a case, it may be reasonable for authors to exclude high risk of bias studies, not because of biased associations, but rather because of inconsistent associations. The review by Cheng et al. exemplifies this concept[20].

In the review addressing the association between smoking and VTE[20], investigators observed substantial differences in point estimates of relative risk from 0.50 to 4.70 with limited overlap in confidence intervals from individual studies (figure 3). They postulated that the association between smoking and VTE would differ when the primary studies were adjusted for other prognostic factors. Indeed, a sensitivity analysis including only adjusted relative risks resulted a much more visually consistent forest plot with the individual study associations ranging from RR 0.90 to 2.87 (figure 3). Dividing the studies into those that adjusted for other prognostic factors resolved the inconsistency issue, producing largely overlapping associations between smoking and VTE.

The pooled effect estimates from unadjusted and adjusted studies did not, however, differ significantly (unadjusted RR: 1.17, 95% CI 1.09 to 1.25; adjusted RR: 1.21, 95% CI 1.15 to 1.26). Therefore, it is reasonable that the authors maintained the unadjusted studies in the primary analysis upon considerations for risk of bias. With regards to inconsistency, however, a separate

meta-analysis for adjusted studies does considerably reduce heterogeneity and thus warrants using adjusted studies meta-analysis without rating down for inconsistency.

Our overall judgment on certainty for the effect of a particular factor may differ for non-contextualised and contextualised settings, even when the same set of studies are examined. In particular, it may be possible that risk of bias may be low for pooled relative effects (i.e. non-contextualised settings), but high for pooled absolute risk predictions for groups defined by a prognostic factor (i.e. contextualised settings). The reason is that estimates of absolute risk predictions are more prone to overfitting concerns; that is, the estimated predicted probabilities that are too close to 0 or 1. This issue is noted in PROBAST[13]. The magnitude and frequency of this risk of bias is uncertain, and currently represents a limitation of our approach to rating the certainty of the evidence in the contextualized framework. Ideally absolute risk predictions conditional on a prognostic factor would address overfitting by using penalisation and shrinkage techniques, and examine calibration in new data; however, this is rarely done. Further consideration of this issue will be given in our planned follow-up work on GRADE for prognostic model research, where the main focus is on absolute risk prediction.

Concluding remarks

Our discussion has demonstrated that the same principles GRADE proposed for body of evidence addressing treatment and overall prognosis work well in assessing bodies of evidence regarding individual prognostic factors, both in non-contextualized and contextualized settings. The goal is ensuring clinical investigators and clinicians understand the certainty associated with the

evidence, and the rationale for certainty ratings: using GRADE guidance, and documenting the logic of its application will achieve that objective.

ACKNOWELDGEMENTS / FUNDING

All authors contributed to the generation of the research hypothesis, participated to the discussion of its content and approved the final version of the manuscript. FF, VZ, AI selected the systematic reviews used as examples and prepared summary of finding tables used in the process. FF drafted the manuscript; AI is the guarantor.

DISCLOSURES

All authors declare they did not receive support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years exist, nor do other relationships or activities that could appear to have influenced the submitted work. [Authors initials] are members of the GRADE working group.

References

1. Riley RD, van der Windt DA, Croft P, et al. Prognosis Research in Health Care: Concepts, Methods, and Impact: Oxford University Press 2019.
2. Riley RD, Hayden JA, Steyerberg EW, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLoS Med* 2013;10(2):e1001380. doi: 10.1371/journal.pmed.1001380 [published Online First: 2013/02/09]
3. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;350:h870. doi: 10.1136/bmj.h870 [published Online First: 2015/03/18]
4. Hultcrantz M, Rind D, Akl EA, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol* 2017;87:4-13. doi: 10.1016/j.jclinepi.2017.05.006 [published Online First: 2017/05/23]
5. Altman DG. Systematic reviews of evaluations of prognostic variables. *BMJ* 2001;323(7306):224-8. [published Online First: 2001/07/28]
6. Hemingway H, Philipson P, Chen R, et al. Evaluating the Quality of Research into a Single Prognostic Biomarker: A Systematic Review and Meta-analysis of 83 Studies of C-Reactive Protein in Stable Coronary Artery Disease. *PLOS Medicine* 2010;7(6):e1000286. doi: 10.1371/journal.pmed.1000286
7. Matino D, Chai-Adisaksopha C, Iorio A. Systematic reviews of prognosis studies: a critical appraisal of five core clinical journals. *Diagnostic and Prognostic Research* 2017;1(1):9. doi: 10.1186/s41512-017-0008-z
8. Riley RD, Moons KGM, Snell KIE, et al. A guide to systematic review and meta-analysis of prognostic factor studies. *BMJ* 2019;364:k4597. doi: 10.1136/bmj.k4597 [published Online First: 2019/02/01]
9. Brożek JL, Akl EA, Jaeschke R, et al. Grading quality of evidence and strength of recommendations in clinical practice guidelines: Part 2 of 3. The GRADE approach to grading quality of evidence about diagnostic tests and strategies. *Allergy* 2009;64(8):1109-16. doi: doi:10.1111/j.1398-9995.2009.02083.x
10. Schunemann HJ, Oxman AD, Brozek J, et al. Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;336(7653):1106-10. doi: 10.1136/bmj.39500.677199.AE [published Online First: 2008/05/17]
11. Hordijk-Trion M, on behalf of the EHSCRI, Lenzen M, et al. Patients enrolled in coronary intervention trials are not representative of patients in clinical practice: results from the Euro Heart Survey on Coronary Revascularization. *European Heart Journal* 2006;27(6):671-78. doi: 10.1093/eurheartj/ehi731
12. Hayden JA, van der Windt DA, Cartwright JL, et al. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;158(4):280-6. doi: 10.7326/0003-4819-158-4-201302190-00009 [published Online First: 2013/02/20]
13. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170(1):51-58. doi: 10.7326/M18-1376 [published Online First: 2019/01/01]
14. Reeves BC, Deeks JJ, Higgins JP, et al. Including non-randomized studies. Higgins JP, Green S, editors Cochrane Handbook for Systematic Reviews of Interventions: The Cochrane Collaboration 2010:391–432.

15. Cook DJ, Fuller HD, Guyatt GH, et al. Risk Factors for Gastrointestinal Bleeding in Critically Ill Patients. *New England Journal of Medicine* 1994;330(6):377-81. doi: 10.1056/nejm199402103300601
16. Kucher N, Kohler HP, Dornhofer T, et al. Accuracy of D-dimer/fibrinogen ratio to predict pulmonary embolism: a prospective diagnostic study. *J Thromb Haemost* 2003;1(4):708-13. [published Online First: 2003/07/23]
17. Perrier A, Roy PM, Sanchez O, et al. Multidetector-row computed tomography in suspected pulmonary embolism. *N Engl J Med* 2005;352(17):1760-8. doi: 10.1056/NEJMoa042905 [published Online First: 2005/04/29]
18. Wells PS, Anderson DR, Bormanis J, et al. Value of assessment of pretest probability of deep-vein thrombosis in clinical management. *Lancet* 1997;350(9094):1795-8. doi: 10.1016/S0140-6736(97)08140-3 [published Online First: 1998/01/15]
19. Sanchez O, Trinquart L, Colombet I, et al. Prognostic value of right ventricular dysfunction in patients with haemodynamically stable pulmonary embolism: a systematic review. *Eur Heart J* 2008;29(12):1569-77. doi: 10.1093/eurheartj/ehn208 [published Online First: 2008/05/23]
20. Cheng YJ, Liu ZH, Yao FJ, et al. Current and former smoking and risk for venous thromboembolism: a systematic review and meta-analysis. *PLoS Med* 2013;10(9):e1001515. doi: 10.1371/journal.pmed.1001515 [published Online First: 2013/09/27]
21. Witlox J, Eurelings LS, de Jonghe JF, et al. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. *JAMA* 2010;304(4):443-51. doi: 10.1001/jama.2010.1013 [published Online First: 2010/07/29]
22. Schandelmaier S, Briel M, Varadhan R, et al. Development of a new Instrument to assess the Credibility of Effect Modification ANalyses (ICEMAN) in randomized controlled trials and meta- analyses. McMaster University, 2019.
23. Foroutan F, Guyatt G, Thabane L, et al. Calculation of absolute risk for important outcomes in patients with and without a prognostic factor of interest. *Journal of Clinical Epidemiology* 2019;In press
24. Foroutan F, Friesen EL, Clark KE, et al. Risk Factors for One-Year Graft Loss Post-Kidney Transplantation. *Clinical Journal of the American Society of Nephrology* 2019;In Press
25. Andre F, McShane LM, Michiels S, et al. Biomarker studies: a call for a comprehensive biomarker study registry. *Nature Reviews Clinical Oncology* 2011;8:171. doi: 10.1038/nrclinonc.2011.4
26. Kyzas PA, Denaxa-Kyza D, Ioannidis JPA. Almost all articles on cancer prognostic markers report statistically significant results. *European Journal of Cancer* 2007;43(17):2559-79. doi: 10.1016/j.ejca.2007.08.030
27. Martin B, Paesmans M, Berghmans T, et al. Role of Bcl-2 as a prognostic factor for survival in lung cancer: a systematic review of the literature with meta-analysis. *British Journal Of Cancer* 2003;89:55. doi: 10.1038/sj.bjc.6601095
28. McShane LM, Altman DG, Sauerbrei W. Identification of Clinically Useful Cancer Prognostic Factors: What Are We Missing? *JNCI: Journal of the National Cancer Institute* 2005;97(14):1023-25. doi: 10.1093/jnci/dji193

29. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50(4):1088-101. [published Online First: 1994/12/01]
30. Debray TPA, Moons KGM, Riley RD. Detecting small-study effects and funnel plot asymmetry in meta-analysis of survival data: A comparison of new and existing tests. *Res Synth Methods* 2018;9(1):41-50. doi: 10.1002/jrsm.1266 [published Online First: 2017/10/05]
31. Peters JL, Sutton AJ, Jones DR, et al. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006;295(6):676-80. doi: 10.1001/jama.295.6.676 [published Online First: 2006/02/10]
32. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-34. [published Online First: 1997/10/06]
33. Vasilevska M, Ross SA, Gesink D, et al. Relative risk of cervical cancer in indigenous women in Australia, Canada, New Zealand, and the United States: a systematic review and meta-analysis. *J Public Health Policy* 2012;33(2):148-64. doi: 10.1057/jphp.2012.8 [published Online First: 2012/03/02]
34. Thompson M, Vodicka TA, Blair PS, et al. Duration of symptoms of respiratory tract infections in children: systematic review. *BMJ* 2013;347:f7027. doi: 10.1136/bmj.f7027 [published Online First: 2013/12/18]

Table 1. Types and goals of prognostic studies

Study type	**Study Goal**	**Examples in the field of atrial fibrillation**
Overall prognosis(#)	Establish the typical risk in a broadly defined population	Risk of bleeding in patients with atrial fibrillation receiving vitamin K antagonists
Prognostic factor	Establish how a particular patient characteristic influences risk	Influence of age on the risk of bleeding in patients with atrial fibrillation
Outcome (or risk) prediction model	Development of a full prognostic model simultaneously considering a number of prognostic factors and classifying patients into various levels of risk	CHADS2 and CHADS-VASC for the risk of stroke HAS-BLED, HEMORRHAGE for the risk of bleeding

(#) – It is equally important to estimate the likelihood of spontaneous resolution of a disease, as discussed in Matthew Thompson et al[34].

Table 2 – Significance of levels of evidence for risk associated with the prognostic factor

Quality level	Definition
High	We are very confident that the variation in risk associated with the prognostic factor (probability of future events in those with/without the prognostic factor) lies close to that of the estimate (#)
Moderate	We are moderately confident that the variation in risk associated with the prognostic factor (probability of future events in those with/without the prognostic factor) is likely to be close to the estimate, but there is a possibility that it is substantially different
Low	Our certainty in the estimate is limited: The variation in risk associated with the prognostic factor (probability of future events in those with/without the prognostic factor) may be substantially different from the estimate
Very low	We have very little certainty in the estimate: The variation in risk associated with the prognostic factor (probability of future events in those with/without the prognostic factor) is likely to be substantially different from the estimate

Prognostic factor studies measure the variation in incidence, i.e. target events over time in a population of interest at risk for the target event, as a function of presenting or not a specific prognostic factor. The target event can be an adverse outcome (e.g. mortality) in patients with a prognostic factor as compared to those without (e.g. BMI > 30 as compared to <25). Sometimes the prognostic factor is the change in some patient characteristic over time (e.g. the combination of the Apgar score at 1 and 5 minutes after delivery).

Figure 1 – Primary (left) and sensitivity (right) analysis for the association between delirium and mortality

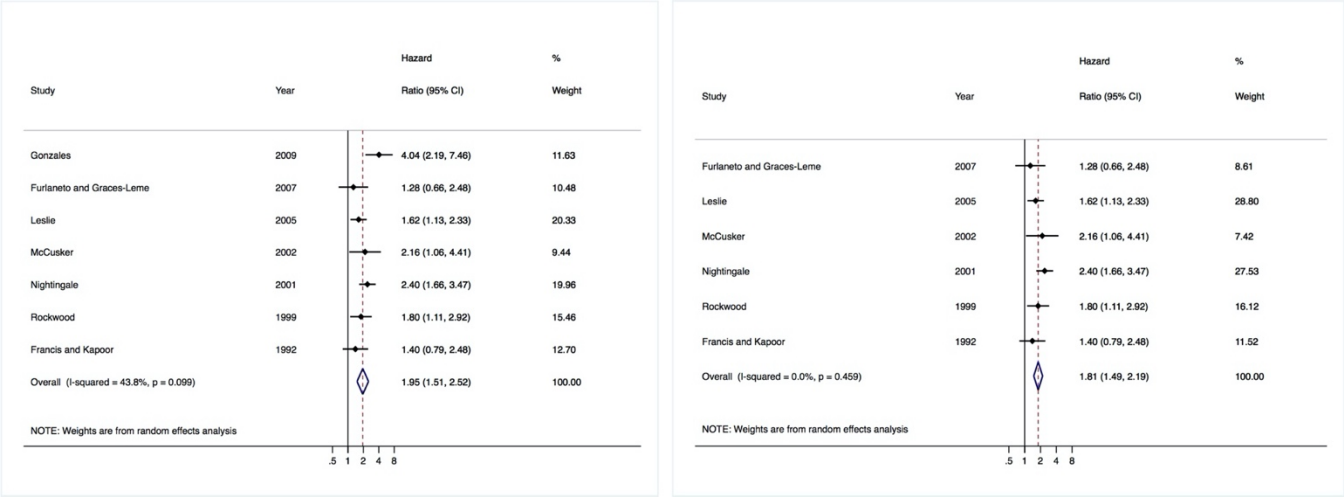


Figure 2 – Funnel plot for visual inspection of publication bias in a review on risk cervical cancer in indigenous versus non-indigenous women.

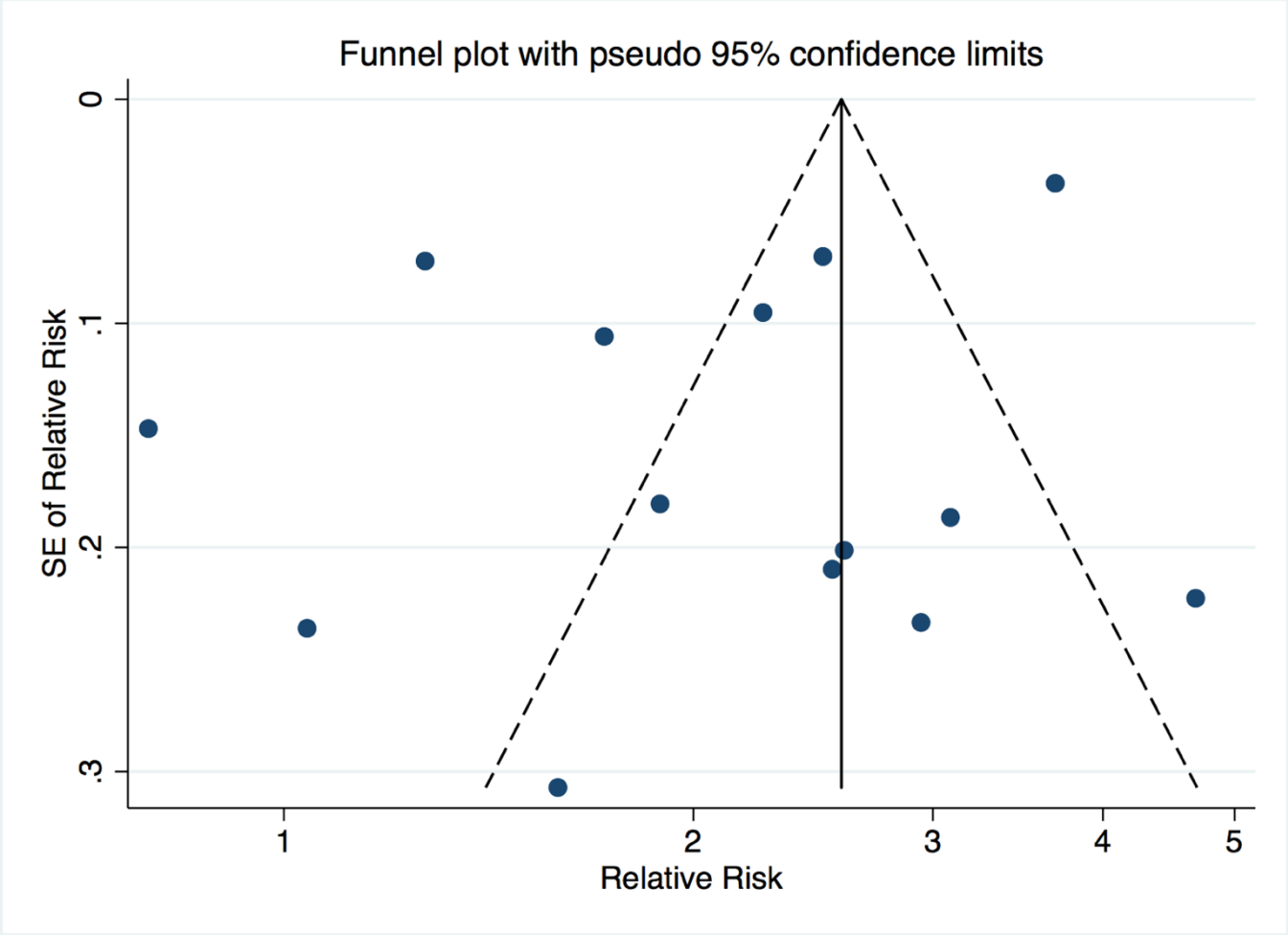
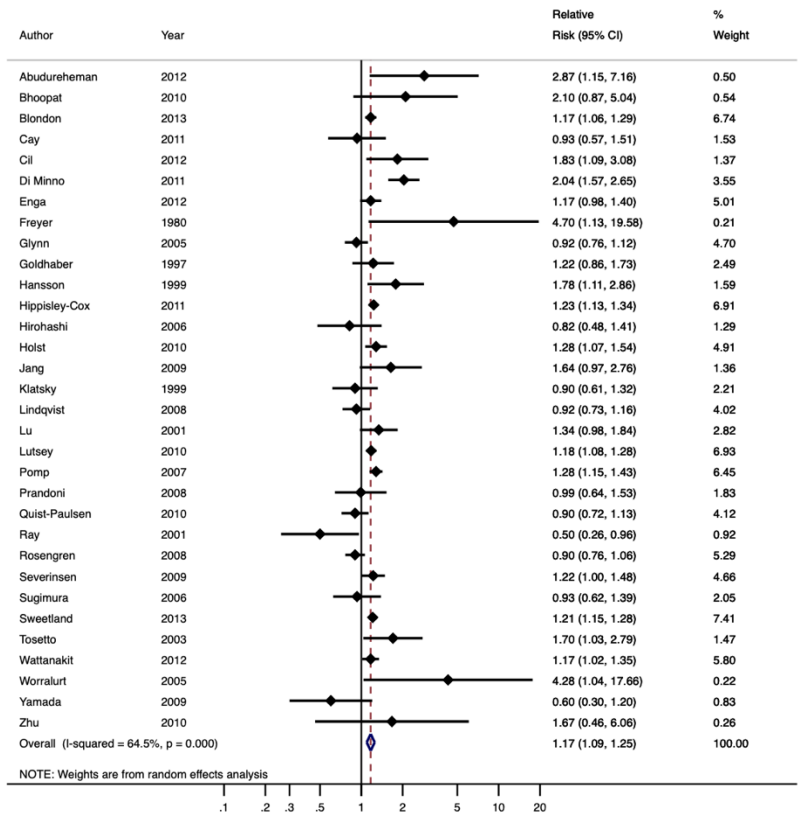
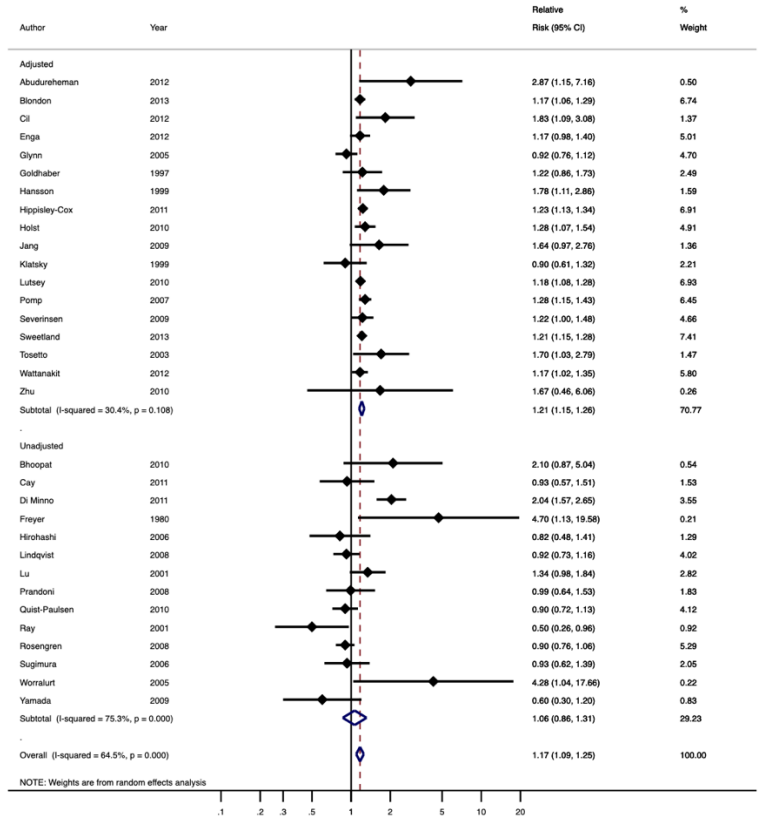


Figure 3 – Primary and subgroup analyses for association between smoking and venous thromboembolism. Subgroup based on adjustment for confounders.





APPENDIX A – Example of a summary of findings table

Recipient Pognostic Factors
Outcome: Graft loss

Prognotic Factors	Study Results and measurements	Absolute effect estimates Baseline	Absolute effect estimates With predictor	Certainty in effect estimates (Quality of evidence)	Plain text summary
Age (10-year increase)	Hazard Ratio 1.16 (CI 95% 1.10 - 1.23) Based on data from 138,824 patients in 12 studies	74 per 1000	85 per 1000 Difference: 11 more per 1000 (CI 95% 7 more - 16 more)	High	Increasing recipient age slightly increases 1-year graft loss
Sex (male vs female)	Hazard Ratio 1.28 (CI 95% 0.98 - 1.67) Based on data from 176,972 patients in 9 studies	63 per 1000	81 per 1000 Difference: 18 more per 1000 (CI 95% 2 fewer - 35 more)	Low Due to seriuos inconsistency, Due to serious publication bias	Recipient sex may have little or no difference on graft loss
Race (white vs others)	Hazard Ratio 0.76 (CI 95% 0.52 - 1.11) Based on data from 169,596 patients in 2 studies	87 per 1000	66 per 1000 Difference: 21 fewer per 1000 (CI 95% 51 fewer - 8 more)	High	Recipient race has little or no difference on graft failure
BMI (1 kg/m2 increase)	Hazard Ratio 1.02 (CI 95% 0.99 - 1.04) Based on data from 51,881 patients in 4 studies	74 per 1000	75 per 1000 Difference: 1 more per 1000 (CI 95% 1 fewer - 3 more)	High	Recipient BMI has little or no impact on graft failure
Dialysis time (per 1-year increase)	Hazard Ratio 1.03 (CI 95% 1.02 - 1.03) Based on data from 51,776 patients in 3 studies	74 per 1000	76 per 1000 Difference: 2 more per 1000 (CI 95% 1 fewer - 2 more)	Moderate Due to serious risk of bias	Years on dialysis proabbly increases graft failure slightly
Diabetes (Yes vs no)	Hazard Ratio 0.99 (CI 95% 0.97 - 1.02) Based on data from 169,015 patients in 2 studies	74 per 1000	73 per 1000 Difference: 1 more per 1000 (CI 95% 2 fewer - 1 more)	High	Recipient diabetes has little or no difference on graft failure
Smoking (ever vs never)	Hazard Ratio 1.59 (CI 95% 1.34 - 1.90) Based on data from 3,156 patients in 2 studies	65 per 1000	104 per 1000 Difference: 39 more per 1000 (CI 95% 23 more - 55 more)	Moderate Due to serious imprecision	Pre-transplant recipient smoking probably increases graft failure slightly
Coronary Artery Disease (Yes vs no)	Hazard Ratio 1.15 (CI 95% 1.03 - 1.27) Based on data from 81,194 patients in 2 studies	73 per 1000	85 per 1000 Difference: 12 more per 1000 (CI 95% 2 more - 20 more)	Moderate Due to serious indirectness	Recipient coronary artery disease probably increases graft failure slightly
Hypertension (Yes vs no)	Hazard Ratio 0.99 (CI 95% 0.98 - 1.01) Based on data from 169,314 patients in 3 studies	74 per 1000	74 per 1000 Difference: 1 fewer per 1000 (CI 95% 2 fewer - 0)	High	Pre-transplant recipient hypertension has little or no difference on graft failure

Chapter 7: Calculation of absolute risk for important outcomes in patients with and without a prognostic factor of interest.

STATUS: Published – Journal of Clinical Epidemiology

https://doi.org/10.1016/j.jclinepi.2019.08.012

Farid Foroutan[1,2], Alfonso Iorio[1,3], Lehana Thabane[1,4], Gordon Guyatt[1,3]

Authors affiliations:

1- Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

2- Ted Rogers Centre for Heart Research, Multi-Organ Transplant Program, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

3- Department of Medicine, McMaster University, Hamilton, Ontario, Canada

4- Biostatistics Unit, St Joseph's Healthcare—Hamilton, Ontario, Canada

ABSTRACT

Background. Primary studies and systematic reviews of prognostic factors commonly analyze and report relative measures of association between the factor(s) and outcome(s) of interest. For decision making, however, guideline panelists, systematic reviewers, and health care professionals at the point of care will ultimately need the absolute risk of the outcome(s) in those with and without the prognostic factor(s) of interest.

Objective. To develop a framework for calculating the absolute risk of the outcome(s) in those with and without the prognostic factor(s) of interest.

Methods. We developed a mathematical approach to calculate the absolute risk of events from the relative measure of association, the total number of events and patients at risk, and the prevalence of the prognostic factor, all of which are usually reported in cohort studies assessing prognostic factors. We demonstrate how simpler approximations lead to biased estimates of absolute risk and thus the need for these formulas. We explain our logical framework using the simplest case, in which the measure of association is a relative risk, and provide extensions of the formula to odds ratios and hazard ratios. The same formulas can be applied to reports providing only the relative measure of association (e.g. case-control studies) by using external evidence regarding prevalence of the prognostic factor and overall risk of events.

Results.

Conclusions. Our proposed formulas facilitate accurate calculation of measures of absolute risk in those with and without prognostic factors of interest for studies reporting the total number of events and patients at risk, the prevalence of the prognostic factor and a relative risk, odds ratio or hazard ratio.

INTRODUCTION

An increasing number of published primary studies and systematic reviews are addressing prognostic factors[1]. These studies identify one or many prognostic factors (e.g characteristic of patients, disease, or result of diagnostic tests) that are associated with a higher or lower likelihood of future patient important events. Some authors differentiate a risk factor (influencing the probability of occurrence of a condition in those currently unaffected) versus a prognostic factor (influencing the occurrence of an outcome in those with an underlying condition such as cancer or cardiovascular disease). In this article, we will use "prognostic factor" for both these situations.

Clarification of the framework for this discussion

The general goal of primary studies or systematic reviews of prognostic factors is to provide robust estimates of the association between one or more prognostic factor(s) and the outcome of interest, for which relative measures– typically relative risk (RR), odds ratio (OR), and hazard ratio (HR) - offer oft-used estimates of magnitude of the association. Authors most often derive these measures of association for variables that are natural dichotomies (e.g those who have suffered a stroke or myocardial infarction versus those who have not) or adopting a threshold for variables that are conceptually continuous, thus creating a dichotomous variable (for instance, a threshold of 65 defines older and younger).

When a variable – such as age – is naturally continuous, one can apply relative measures of association to a range of the continuous variable (e.g. the increase in odds of an event per decade of older age). When interested in the absolute risk of an outcome for a specific patient with a certain age, one may consider evidence from an applicable cohort. This cohort will have an

average age and an overall risk of the outcome, the latter representing the absolute risk for individuals with the average age. One can apply the relative effect for age and the difference in age of the specific patient from the average to the overall risk of the outcome to calculate the absolute risk for an individual.

Dichotomies, however, require additional considerations that make the calculations more challenging. In the rest of this paper we will discuss calculation of absolute risk within the context of natural dichotomies, or continuous variables that investigators have chosen to treat and report as dichotomies.

For any dichotomous prognostic factor there are two levels of that factor – for instance, consider the case for history of stroke. The natural way of expressing these two levels is as those with and without a prognostic factor that increases risk – for instance, one could consider individuals with a previous stroke as those with the prognostic factor of interest, and individuals without a previous stroke as those without the prognostic factor of interest. In our discussion, we will use the language in this way.

When authors provide a single measure of association for a particular prognostic factor, they assume that the relative difference in prognosis between those with and without that prognostic factor is constant across different levels of risk determined by other prognostic factors. For instance, the assumption would be that having had a previous stroke has the same increase in relative impact on risk of cardiovascular death (i.e. odds ratio, relative risk, or hazard ratio) whether or not one is diabetic, or hypertensive, or has suffered a prior myocardial infarction. The extent to which prognostic factors are similar across patient groups (that is, the extent to which interactions between prognostic factors exist) is currently uncertain. The absolute risk of

cardiovascular mortality will, however, differ depending on the patient status with respect to other prognostic factors (e.g. those who do not have diabetes, hypertension or a prior myocardial infarction will have a lower risk than those who do).

Need for absolute measures of risk for decision making

Patients and clinicians make evidence based medical decisions by comparing the expected likelihood of patient important outcomes across alternative management options. In making such decisions, it is often useful to consider a particular probability of the outcome of interest as a decision threshold. For instance, in patients with a prior deep venous thrombosis (DVT) one might consider the reduction in recurrent thrombosis that would make one willing to undergo the burdens and bleeding risk associated with anticoagulation. That threshold in turn depends on the values and preferences associated with benefits and harms/burdens of the intervention.

Optimal use of such decision thresholds requires absolute measures of risk. For instance, consider a patient with a first episode of unprovoked venous thromboembolism, at the end of a 6 months treatment course, facing the decision whether or not to continue treatment with a prophylactic anticoagulant. In this situation, clinicians may utilize a 5% risk of recurrence at 1 year as the threshold that warrants continued prophylactic anticoagulation[2]. A recent cohort study[3] showed that the risk of recurrence at 1-year after treatment in patients with negative D-dimer (measured at the end of the treatment) is 7.4%. Male patients, who represent 50.3% of the cohort, are at a 2.2 times greater risk of recurrence compared to women. One may intuit that stopping treatment in women is safe (if the relative risk of men to woman for recurrence is 2.2, the risk of recurrence in woman must be below 5%). The decision to stop, indeed, requires knowledge of the absolute risk of recurrence in women to be used in conjunction with their

values and preferences (their agreement with the 5% threshold for stopping or continuing anticoagulation).

This example provides one illustration of the principle that guideline panelists and practicing physicians require absolute measures of risk to make use of a prognostic factor in clinical decision making, and in particular where patients lie in relation to a threshold. If, as is often the case, systematic reviews or primary studies of prognostic factors report only relative measures of association, arriving at measures of absolute effect may become a daunting challenge.

Challenges with transforming relative effects into absolute measures of risks

Continuing with the anticoagulant example, one may be tempted to calculate the risk in woman by dividing 7.4% (the risk in the overall cohort) by 2.2 (the relative risk of men compared to woman)[3]. With this approach, one would conclude an absolute risk of 3.4% at 1-year after treatment in woman with negative D-dimer test at the end of their treatment, less than the 5% threshold for stopping prophylactic therapy. The authors, however, report recurrence rates of 9.5% for men and 5.3% for women[3], suggesting the opposite clinical decision: continue treatment with anticoagulation.

This example makes evident that calculating the absolute risk for the patients with or without a prognostic factor, needed to decide on where patients lie with respect to a management threshold, is more complicated than applying the relative measure of association to the risk of events in the entire cohort. Unbiased calculations require us to consider the prevalence of the prognostic factor.

Aim of the paper

In this paper, we provide a simple framework for calculation of absolute risks for patients with and without a prognostic factor starting from relative measures of risk, the prevalence of a prognostic factor in the population, and the risk of the patient important outcome in the population of interest.

Calculation of absolute risk when not explicitly reported

To understand the process of obtaining absolute risks when authors fail to report explicitly, we will consider an observational study from our group assessing the association between primary graft dysfunction (comorbidity that occurs within the first 24 hours after heart transplantation) and all-cause 1-year mortality[4]. Although we did report explicitly on subgroup risk, the data works well for illustrative purposes.

This study reported on a retrospective cohort of 412 adult heart transplant recipients, of whom 17% developed primary graft dysfunction (PGD). Therefore, our prognostic factor, PGD, has a prevalence of 0.17. The relative risk for 1-year mortality associated with primary graft dysfunction (PGD) versus no PGD was 4.88. The overall risk of all-cause 1-year mortality (risk of outcome irrespective of whether patients have PGD or not) was 11.2% (46 cases of death / 412 total cohort). To calculate the absolute risk in patients with PGD, one might be tempted to multiply the overall risk by the relative risk of 4.88 which equals 55%. It turns out, however, that this represents an overestimate of the true absolute risk.

The relative risk of 4.88 provides the best estimate of the risk of mortality in patients with PGD (π_1), compared to those without PGD (π_1). Thus, the relative risk can be presented as:

$$\frac{Risk\ of\ mortality_{PGD}}{Risk\ of\ mortality_{No\ PGD}} = RR_{PGD}$$

$$\frac{\pi_1}{\pi_2} = 4.88 \text{ (Equation 1)}$$

The risk of mortality in the entire cohort is the combined risk of this outcome in those with and without PGD. We can present the risk in the entire cohort as:

$$Risk\ of\ mortality_{entire\ cohort} = prevalence_{PGD} \times risk\ of\ mortality_{PGD} + prevalence_{No\ PGD} \times risk\ of\ mortality_{No\ PGD}$$

$$0.112 = 0.17 \times \pi_1 + 0.83 \times \pi_2 \text{ (Equation 2).}$$

We can re-arrange Equation 1 to:

$$\pi_1 = 4.88 \times \pi_2 \text{ (Equation 3).}$$

Now we can substitute equation 3 into equation 2:

$$0.112 = 0.17 \times (4.88 \times \pi_2) + 0.83 \times \pi_2 \text{ (Equation 4).}$$

After the substitution, we can solve equation 4 for the value of y, which represents the absolute risk of mortality in those without PGD:

$$0.112 = 0.8296 \times \pi_2 + 0.83 \times \pi_2$$

$$0.112 = 1.6596 \times \pi_2$$

$$\pi_2 = \frac{0.112}{1.6596}$$

$$\pi_2 = 0.07$$

Therefore, the absolute risk of mortality in patients without PGD is 7%. Substituting this value in equation 3 solves the absolute risk of mortality in those with PGD:

$$\pi_1 = 4.88 \times 0.07$$

$$\pi_1 = 0.33$$

Thus, the absolute risk of mortality in those with PGD is 33%. This is, indeed the absolute risk from the primary data.

One can extend the approach for application to hazard ratios and odd ratios. Equation 2 showed that the overall risk of the outcome is a combination of the absolute risk in those with and without the prognostic factor of interest. Calculating the absolute risk requires considering the proportion of patients with the prognostic factor.

Again, we denote the absolute risk of mortality in patients with PGD as π_1, and the risk in those without PGD as π_2. We can consider a ratio between the overall risk of the outcome and the risk of outcome in those with our prognostic factor of interest, PGD.

$$\frac{risk\ of\ mortality_{entire\ cohort}}{risk\ of\ mortality_{PGD}}$$

$$\frac{prevalence_{PGD} \times risk\ of\ mortality_{PGD} + \ prevalence_{\ No\ PGD} \times risk\ of\ mortality_{\ No\ PGD}}{risk\ of\ mortality_{PGD}}$$

$$\frac{0.17\ \pi_1 + 0.83\ \pi_2}{\pi_1} \text{ (Equation 5)}$$

If we divide the numerator and denominator of the equation 5 by the risk of mortality in patients without PGD (y), we obtain the following equation:

$$\frac{0.17\frac{\pi_1}{\pi_2} + 0.83\frac{\pi_2}{\pi_2}}{\frac{\pi_1}{\pi_2}} \text{ (Equation 6)}$$

Equation 6 can be re-written as:

$$\frac{0.17\ RR_{PGD} + 0.83}{RR_{PGD}}$$

The ratio calculated from this equation will necessarily be the same as the ratio between the risk of mortality in the entire cohort and the risk of mortality in patients with PGD.

An equality statement

$\pi_1 = risk\ in\ patients\ with\ prognostic\ factor$

$\pi_2 = risk\ in\ patients\ without\ prognostic\ factor$

$p_1 = prevalence\ of\ patients\ with\ prognostic\ factor$

$p_2 = prevalence\ of\ patients\ without\ prognostic\ factor$

$RR = relative\ risk\ = \frac{\pi_1}{\pi_2}$

$Overall\ Risk = risk\ of\ the\ outcome\ in\ the\ entire\ cohort$

Rule:

$$\frac{Overall\ Risk}{\pi_1} = \frac{(p_1 \times\ RR) +\ p_2}{RR}$$

Proof:

$$\frac{Overall\ Risk}{\pi_1}$$

$$= \frac{\pi_1 \times p_1 + \pi_2 \times p_2}{\pi_1}$$

$$= \frac{\frac{\pi_1}{\pi_2} \times p_1 + \frac{\pi_2}{\pi_2} \times p_2}{\frac{\pi_1}{\pi_2}}$$

$$= \frac{RR \times p_1 + p_2}{RR}$$

Expansion of the principles to associations depicted with hazard ratio and odds ratio

The expansion of the calculation for associations depicted with hazard ratio and odds ratio requires modifications. The ratio calculated on the left side of our equality statement is the same as the ratio between overall odds and odds of the outcome in the group with the factor of interest. For a hazard ratio, the ratio calculated on the left side of our equality statement is the same as the ratio between overall hazard rate and the hazard rate in the group with the factor of interest. The derivation of odds and hazard requires transformation of probability. To

transform probability to odds, we need to divide the probability of having the outcome by the probability of not having the outcome.

$$odds = \frac{p}{1-p} \text{ (Equation 7)}$$

To transform probability to hazard, we need to obtain the probability of not having the outcome (1 – probability of having the outcome; often referred to as survival probability). By taking the natural logarithm of the probability of not having the outcome, dividing by -1 and the time, we obtain the hazard.

$$hazard\ (h) = \frac{\ln(1-p)}{-1 \times time} \text{ (Equation 8)}$$

$$hazard\ (h) = \frac{\ln(survival)}{-1 \times time} \text{ (Equation 8)}$$

Example with hazard ratio

We can refer back to our example looking at the association between PGD and 1-year mortality. The Cox regression model showed that presence of PGD is associated with a 6-fold increase in hazard of mortality (HR 6.14, 95% CI 3.44 to 10.96). The overall 1-year survival of the cohort of transplant recipients was 89%. The prevalence of PGD was 17%. Using equation 6, we can derive the expression of 0.17 HR + 0.83. Using equation 8, we can convert the overall survival to an overall hazard rate of 0.116. Using these values, we can set-up an equality equation as seen above.

$$\frac{0.17HR + 0.83}{HR_{PGD}} = \frac{overall\ hazard}{hazard_{PGD}}$$

$$\frac{1.87}{6.14} = \frac{0.116}{hazard_{PGD}}$$

$$hazard_{PGD} = 0.381$$

This hazard can be converted back to a 1-year survival probability by using the inverse of equation 8.

$$survival = e^{(-1\times1\times0.381)}$$

$$survival = 68\%$$

This is in fact, the actual one-year survival we observed in patients with PGD. The hazard ratio for not having PGD is 0.16 (the inverse of the hazard ratio for PGD). We can apply this hazard ratio to the survival probability calculated for patients with PGD.

$$survival_{No\ PGD} = 0.68^{0.16}$$

$$survival_{No\ PGD} = 0.93$$

We can now use the one-year probability of survival in the presence of PGD to inform a clinical encounter with heart transplant recipients presenting PGD. Prior to transplantation we may have quoted their risk of 1-year mortality at 11%. When, unfortunately, the transplant recipient develops PGD within the first 24 hours post-transplant, we can – regretfully - inform them that their mortality risk has increased to 32%. This is certainly more informative than telling the patient that his HR of death is now 6.14 times higher compared to someone who doesn't have PGD, and more accurate than reporting an increase in risk from 11% to 67% (6.14 times 11).

In another example one may consider the prevention of recurrent VTE after completing a full initial course of anticoagulation. Risk of recurrence is higher in males than females (HR 1.79, 95% CI 1.33 – 2.43)[5]. A possible threshold to withhold treatment is a recurrence rate of 5% or less at 1 year[2 6]. Does knowing the hazard ratio for male sex help in deciding if management may reasonably differ between males and females? It does not because, applying our approach, the

absolute recurrence rates in males (9.55, 95% CI 7.4 to 11.4) and females (5.3%, 95% 4.2 to 6.7) are both above the 5% threshold[3], a result consistent with that reported in another study[7].

To facilitate these calculations, we devised an online calculator based on the aforementioned formula: http://hiru.mcmaster.ca/AbsoluteRiskCalculator/. Using this online calculator, within the appendices, we provide a number of worked case scenarios.

DISCUSSION

In this manuscript we provide a framework for the calculation of absolute measures of risk, needed for decision making, from relative measures (RR, OR, HR), usually reported in published studies or systematic reviews. Authors of primary studies and systematic reviews should consider directly reporting absolute measures of risk in patients with and without risk factors, or at least reporting the observed prevalence of prognostic factor(s) and the overall risk of outcome(s) in the population they studied, the information required to calculate absolute risks. When this information is available, guideline panelists, health care professionals and policy makers will be able to use our proposed formulas to calculate the absolute measures they need.

Specifically, we show, and provide proof, that with the use of a prognostic factor prevalence, overall risk of outcome, and the effect estimate of the prognostic factor, one can develop an equality statement that allows solving for the absolute risk of the outcome in those with and without the prognostic factor.

As we mentioned earlier, one may be tempted to apply the relative effect of a prognostic factor to the overall risk of an outcome within a certain cohort. In almost all instances, this approach, which does not consider the prevalence of the prognostic factor in the population, will result in an overestimation of the absolute risk of outcome (55% risk of mortality with PGD when

we don't consider prevalence versus 33% risk of mortality with PGD when we do consider prevalence).

One may apply our proposed calculations directly to bodies of evidence providing the three needed components, i.e. prevalence of the prognostic factor(s), relative measure of risk association(s) and overal risk in the underlying population (all individuals, irrespective of them having or not the prognostic factor of interest). In the context of systematic reviews, the most conservative approach would be to calculate the median prevalence and total risk of the outcome as reported by the individual studies[8].

Authors may, however, not report the prevalence of the prognostic factor and the overall risk of the outcome for the population under study. When this is the case, one may reasonably rely on indirect evidence from other studies reporting the prevalence of the prognostic factor and overall risk of outcome to conduct the necessary calculations.

Conclusion

Authors of prognostic factor studies may fail to report absolute risks for the outcome of interest in those with and without the factor of interest and authors of systematic reviews of prognostic factors may thus find the information unavailable. Provided the overall risk of the outcome and the prevalence of the prognostic factor are reported or known, however, authors of reviews and guideline panelists may calculate absolute risks for each group separately. Such calculations contextualize the relative effect estimates for prognostic factors, allowing for a more informed decision-making process. Authors of future systematic reviews would well serve their clinician readers by conducting the proposed calculations, or at least summarizing and reporting all the three components needed for the calculation, so that guideline panelist, practicing

clinicians or health policy makers may use our framework to calculate the relevant absolute risks at the decision-making point.

ACKNOWELDGEMENTS / FUNDING

Farid Forotuan and Alfonso Iorio devised the idea. Farid, Alfonso, Lehana, and Gordon authored the manuscript and made necessary revisions. Lehana Thabane verified the mathematical proof and calculations.

DISCLOSURES

None of the authors have any conflict of interest to declare

REFERENCES

1. Matino D, Chai-Adisaksopha C, Iorio A. Systematic reviews of prognosis studies: a critical appraisal of five core clinical journals. *Diagnostic and Prognostic Research* 2017;1(1):9. doi: 10.1186/s41512-017-0008-z
2. Kearon C, Iorio A, Palareti G, et al. Risk of recurrent venous thromboembolism after stopping treatment in cohort studies: recommendation for acceptable rates and standardized reporting. *J Thromb Haemost* 2010;8(10):2313-5. doi: 10.1111/j.1538-7836.2010.03991.x [published Online First: 2010/08/27]
3. Douketis J, Tosetto A, Marcucci M, et al. Risk of recurrence after venous thromboembolism in men and women: patient level meta-analysis. *BMJ* 2011;342:d813. doi: 10.1136/bmj.d813 [published Online First: 2011/02/26]
4. Foroutan F, Alba AC, Stein M, et al. Validation of The International Society of Heart and Lung Transplantation (ISHLT) Primary Graft Dysfunction instrument in heart transplantation. *Journal of Heart and Lung Transplantation* 2019;In Press
5. Douketis J, Tosetto A, Marcucci M, et al. Patient-level meta-analysis: effect of measurement timing, threshold, and patient age on ability of D-dimer testing to assess recurrence risk after unprovoked venous thromboembolism. *Ann Intern Med* 2010;153(8):523-31. doi: 10.7326/0003-4819-153-8-201010190-00009 [published Online First: 2010/10/20]
6. Iorio A, Kearon C, Filippucci E, et al. Risk of recurrence after a first episode of symptomatic venous thromboembolism provoked by a transient risk factor: a systematic review. *Arch Intern Med* 2010;170(19):1710-6. doi: 10.1001/archinternmed.2010.367 [published Online First: 2010/10/27]
7. BAGLIN T, DOUKETIS J, TOSETTO A, et al. Does the clinical presentation and extent of venous thrombosis predict likelihood and type of recurrence? A patient-level meta-analysis. *Journal of Thrombosis and Haemostasis* 2010;8(11):2436-42. doi: 10.1111/j.1538-7836.2010.04022.x
8. Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol* 2013;66(2):158-72. doi: 10.1016/j.jclinepi.2012.01.012 [published Online First: 2012/05/23]

Appendix A – Case scenario for calculation of absolute risks

Wiltox et al. studied the association between delirium and mortality in "Delirium in elderly patients and the risk of post discharge mortality, institutionalization, and dementia: a meta-analysis". From the supplementary material of the systematic review, we can calculate a median prevalence of 31% for delirium. The average risk of mortality (i.e. the risk in the entire population, irrespective of the patients having or not the prognostic factor) amongst the 7 studies is 185 per 1000 persons. The relative risk of mortality in persons with delirium as compared to those without is 1.55 (95% CI 1.17 to 1.89). Using our calculator, we can calculate the absolute risk in patients with delirium (equal to 245/1000). The risk in those without delirium will be 158/1000.

Witlox J, Eurelings LS, de Jonghe JF, et al. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. JAMA 2010;304(4):443-51. doi: 10.1001/jama.2010.1013 [published Online First: 2010/07/29]

Appendix B – Examples from a systematic review and meta-analysis

The following examples are from the following systematic review and meta-analysis: Foroutan F, Clark K, Malik A, et al. Predictors of Mortality Post Lung Transplantation: Systematic Review and Meta-Analysis. The Journal of Heart and Lung Transplantation 2019;38(4):S333. doi: 10.1016/j.healun.2019.01.841

Predictor	Prevalence	Study Results and measurements	Absolute effect estimates	
			Baseline	With predictor
Donor factor				
Donor Sex (Male vs Female)	60%	Hazard Ratio 0.92 (CI 95% 0.88 - 0.98) Based on data from 25,200 patients in 2 studies	165 per 1000	153 per 1000
			Difference: 12 fewer per 1000 (CI 95% 19 fewer - 5 fewer)	
Transplant factor				
Type of Tx (BLTx vs SLTx)	57%	Hazard Ratio 0.81 (CI 95% 0.75 - 0.87) Based on data from 32,456 patients in 5 studies	176 per 1000	145 per 1000
			Difference: 31 fewer per 1000 (CI 95% 42 fewer - 21 fewer)	
CMV Status (R -, D +)	16%	Hazard Ratio 1.26 (CI 95% 1.11 - 1.44) Based on data from 25,200 patients in 2 studies	152 per 1000	188 per 1000
			Difference: 36 more per 1000 (CI 95% 15 more - 58 more)	
CMV Status (R +, D +)	21%	Hazard Ratio 1.09 (CI 95% 1.02 - 1.17) Based on data from 25,200 patients in 2 studies	155 per 1000	168 per 1000
			Difference: 13 more per 1000 (CI 95% 3 more - 24 more)	
CPB use (Yes vs No)	22%	Hazard Ratio 1.31 (CI 95% 1.03 - 1.68) Based on data from 2,166 patients in 5 studies	149 per 1000	190 per 1000
			Difference: 41 more per 1000 (CI 95% 4 more - 82 more)	
Recipient factor				
Recipient Age (per 10 years)	N/A	Hazard Ratio 1.16 (CI 95% 1.13 - 1.19) Based on data from 25,124 patients in 4 studies	158 per 1000	181 per 1000
			Difference: 23 more per 1000 (CI 95% 19 more - 27 more)	
Recipient Age (<20 vs >20)	4%	Hazard Ratio 1.37 (CI 95% 1.16 - 1.60) Based on data from 754 patients in 2 studies	156 per 1000	207 per 1000
			Difference: 51 more per 1000 (CI 95% 23 more - 81 more)	
Recipient hypertension (Yes vs No)	17%	Hazard Ratio 1.34 (CI 95% 1.04 - 1.73) Based on data from 1,104 patients in 2 studies	150 per 1000	196 per 1000
			Difference: 46 more per 1000 (CI 95% 6 more - 90 more)	
BMI (<18 vs 18 - 23)	15%	Hazard Ratio 1.30 (CI 95% 1.14 - 1.49) Based on data from 10,370 patients in 2 studies	152 per 1000	193 per 1000
			Difference: 41 more per 1000 (CI 95% 20 more - 64 more)	
Aetiology (Obstructive vs Restrictive)	37%	Hazard Ratio 0.85 (CI 95% 0.78 - 0.92) Based on data from 24,384 patients in 2 studies	167 per 1000	143 per 1000
			Difference: 24 fewer per 1000 (CI 95% 35 fewer - 12 fewer)	
Coronary Artery Disease (Yes vs No)	16%	Hazard Ratio 1.58 (CI 95% 1.13 - 2.22) Based on data from 947 patients in 3 studies	146 per 1000	220 per 1000
			Difference: 73 more per 1000 (CI 95% 18 more - 137 more)	

Chapter 8: Generic Rating of Allograft Function post-Transplant: GRAFT – Heart

STATUS: Unpublished

Farid Foroutan[1,3], Maureen Meade[3], Ana Carolina Alba[1], Tieghan Killackey[1], Gordon Guyatt[3]

1- Ted Rogers Centre for Heart Research, Multi-Organ Transplant Program, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada

2- Department of Medicine, University of Toronto, Toronto, Ontario, Canada

3- Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada

ABSTRACT

Introduction

Deceased donors provide many organs for transplant to individuals with failing hearts, livers, kidneys and other organs. Optimal care of deceased donors in the intensive care units may increase the number of available organs, and promote their longevity. Research informing the care of deceased organ donors poses unique challenges, including the possibility that 1 deceased donor will provide up to 8 organs for transplantation. An attractive outcome measure that could enhance the feasibility of randomized trials in deceased donor management would measure post-transplant organ function. However, 'function' at any point in time is organ-specific. Therefore, we developed an outcome measure that integrates the function of multiple organs early on post-transplant.

Methods

We held regular meetings with methodology, biostatistics, and clinical experts to develop the cardiac version of our measurement instrument. Through regular meetings and presentations, we ensured face validity through iterative feedback. To optimize reliability and accuracy, our definition for each classification category is based on patient prognosis. Using the best available evidence, we developed specific thresholds and guides for classification of patients. We conducted a mixed method user testing to assess inter-rater reliability and usability.

Results

Our team developed a 6-point generic rating instrument for graft function to be applied post-transplant across all major transplantable organs. GRAFT is designed to applied at the time of discharge, 1-month post-transplant, or at the time of death (whichever occurs first). We classify function as normal, impaired but likely to gain normal function, impaired and unlikely to gain

normal function, severely impaired but likely to gain some function, severely impaired and unlikely to gain some function, and irreversible graft failure. We observed acceptable reliability (kappa of 0.87, 95% CI 0.62 – 1.00) and acceptable usability with a system usability scale score of 75 (range of 72.5 – 92.5). Participants found GRAFT easy to use and useful as an outcome measure. We revised the current version of our GRAFT instrument based on the feedback we received from participant's.

Conclusion

In this study, we developed and evaluated the cardiac version of GRAFT, ensuring face validity and demonstrating reliability and usability. The development of our tool will better facilitate conduct of future research to improve care of deceased organ donors and increase the availability and function of donated organs.

INTRODUCTION

One important strategy to improve both the quantity and quality of organs for transplantation is to improve the medical care of deceased organ donors[1-3]. Donation science is a relatively new field fraught with unique challenges. These challenges largely explain the current lack of research to inform donor care[3].

One specific challenge is identifying a primary outcome for clinical trials: one that occurs frequently enough, and over a suitable time frame, to ensure feasibility. Although counting the number of organs transplanted is important, studying the downstream effects of donor care on transplant recipients will be critical. Graft survival is a critical outcome, but achieving complete follow-up over the long-term presents formidable resource and logistic challenges. A measure of early post-transplant organ function represents an attractive intermediate outcome that could enhance the relevance and feasibility of clinical trials of deceased donor care.

Because one deceased donor may provide up to 8 organs, researchers will apply a study intervention in donors and measure outcomes among the recipients of a number of organ types. This raises further challenges to integrating measures of dys/function, across organs (e.g. ejection fractions, glomerular filtration rates, liver biochemistry, arterial oxygenation, and glucose tolerance). An outcome that captures the function of multiple organs, and is likely to reflect longer-term organ function and survival, would enhance the feasibility of potentially practice-changing clinical trials. A generic outcome will also increase the statistical power of these trials. In addition to providing greater statistical power in future studies of deceased donor care, such an instrument might also facilitate the assessment and performance of groups or institutions involved in organ donation and recovery. Ultimately, the instrument may provide a critical measure of graft function that facilitates communication across clinicians and researchers

involved in organ donation. To this end, the objective of this study was to develop an instrument for generic rating of allograft function specifically for clinical trials in deceased donor management. We named our instrument GRAFT (Generic Rating of Allograft Function post Transplant). In this paper, we modelled the methods and development for heart transplantation. The same basic methods will be used for all other solid organs.

METHODS

Development of the instrument

For the development of the instrument, we created a core team of methodology experts (Drs. Gordon Guyatt and Steve Hanna), with experience in development of measurement instruments or biostatistics, and clinicians with expertise in deceased donor management (Dr. Maureen Meade) or post-transplant care of recipients (Drs. Ana Carolina Alba for heart, Christine Ribic and Darin Treleaven for kidney, Aman Sidhu for lung, and Zita Galvin for liver). Beginning in September 2016, we met regularly for a total of 30 meetings. We dedicated the earlier meetings to determining the measurement domains and classification categories. Specifically, we discussed the timing of GRAFT application, categories of graft function/dysfunction, and their definitions. Once the core group agreed on the first conceptualization of the GRAFT instrument, we presented the preliminary version at two interprofessional rounds (cardiology rounds and multi-organ transplant rounds) and one national conference (Canadian Society of Transplantation Conference). These presentations provided feedback regarding the face validity of our definitions for each classification category. We presented our instrument, at a cardiology round, to transplant physicians, surgeons, and fellows in training and asked them to assess each measurement domain and the applicability of our instrument.

Each meeting and each presentation provided an opportunity for feedback, resulting in iterative changes and improvements to the instrument (figure 1 and 2).

Evidence-based Guides

Based on the definitions of each category in the GRAFT instrument (figure 2), we created evidence-based thresholds for classification by linking graft function/dysfunction heart transplant recipients to their prognosis within the first year-post-transplant. We sought systematic reviews and meta-analyses of prognostic factors for 1-year mortality (methods and results presented in Foroutan et al. BMJ Heart. 2018[8]). Informed by this review, and other sources of evidence on graft function[9], the core group came to a consensus on optimal thresholds for classification of patients in each category.

User Testing

To explore the experience of health care professionals using the cardiac version of the GRAFT instrument in the post-transplant setting, we conducted a mixed methods user testing study. To conduct this study, we sought and received approval from the research ethics board of University Health Network.

Participant selection

Because, ultimately, heart transplant cardiologists will be applying the instrument during clinical care of patients in clinical trials, we included a convenience sample from Toronto General Hospital to participate in the user testing process. We defined heart transplant cardiologists as cardiologist involved in the primary of care of heart transplant recipients (either as faculty or fellows in training). We contacted all participant physicians through email correspondence and/or individual conversations. Participating physicians provided informed consent prior to enrollment.

Testing procedure

Trained research staff conducted all user testing in a private setting convenient for the participant. We first provided each participant with a brief overview of the testing procedures. Testing included 10 paper-based clinical scenarios with varying levels of graft function/dysfunction post-transplant, and a paper version of the GRAFT-Cardiac instrument. We developed 10 characteristic heart transplant patient scenarios that summarize the experience of 10 real patients and that provide the necessary information for the classification of patients using our GRAFT instrument. To ensure a wide breadth of characteristic heart transplant scenarios, we selected patients that ranged from challenging heart transplant cases with complicated post-operative course to normal cases with uncomplicated hospital discharge. Specifically, each scenario provided demographic information (recipient age, sex, and prior listing status), pre-transplant medications, pre-transplant comorbidities (such as hypertension, hypercholesteremia, pre-transplant mechanical circulatory support [MCS], duration of MCS, and pre-transplant sensitization of human leukocyte antigens [HLA]). In each case, we also provided users with information on intra-operative and post-operative events/complications, and the duration of cold-ischemic time. We also provided all measures necessary for classification of graft function using the GRAFT instrument.

We instructed users to use the GRAFT instrument as in a real-life setting of post-transplant patient care. Each testing session took no more than 60 minutes. The iterative feedback obtained from participants provided guidance when making modifications to the instrument.

We utilized the concurrent Think Aloud[10] approach and qualitative assessment as self-report methods for documenting end users experience as they first encountered the GRAFT instrument[11-13]. We encouraged participants to verbalize thoughts, expectations, observations

and feelings about the instrument as they progressed through the 10 scenarios applying the GRAFT-Cardiac instrument. The interviewer prompted participants to immediately elaborate on their comments, or made a note to probe at the end of the session. They fell silent for more than a few seconds, the interviewer provided reminders to users to think aloud. For example: "Please continue to tell me aloud what you are thinking as you work out your rating of this patient?" or "Can you describe what you are considering about the patient at the moment?" or "Are you having doubts or uncertainties about the patient scenario, or how you will be rating the patient using the GRAFT instrument?". Subsequently, we continued the interview and asked participants about the structure and content of the instrument including wording, clarity, instructions, ease of use, limitations, and possible improvements; and reflections on the user testing process itself. Through these, we detected usability challenges and solicited further ideas to improve the instrument. The semi-structured interviews lasted approximately 30 minutes (See Appendix A for interview guide).

At the end of the session, we asked respondents to complete the System Usability Scale (SUS) to provide subjective overall feedback on usability and their satisfaction. The SUS is a validated, highly reliable, 10-item survey that uses a 5-point Likert scale from 1 (Strongly disagree) to 5 (Strongly agree)[14]. SUS rating categories are as follows: 0–64 is unacceptable, 65–84 is acceptable, 85–100 is excellent. A score of 82 represents the likelihood to recommend (LTR) threshold Promoters (i.e. people likely to recommend the system), whereas Detractors (i.e. people who would not recommend the system) have an average SUS score of 67.25[14].

We audio recorded and transcribed all interviews.

Data analysis

We used deductive content analysis to analyze the transcribed audio-recordings of the qualitative interviews. We developed an analysis matrix that categorized participants' feedback by content area15. We adapted our content areas from Morville's honeycomb model, a tool that captures facets of the user experience: usability, usefulness, desirability, findability, accessible, credibility, and value16. We adapted the facets that are most useful for the assessment of the GRAFT instrument: credibility, usefulness, accessibility, findability, and usability. We define each of these as:

Credible: The information used at the time of classification are trustworthy for prognosticating the risk of 1-year mortality post heart transplantation.

Useful: The classification of the patients into each of the four categories is useful for clinicians involved in research and those caring for transplant recipients in the follow-up period.

Accessible: The projection of graft recovery may be highly dependent on the years of experience incurred by the participant. The guidance for response to this aspect of the instrument is ideally accessible enough to cardiologists with any level of experience.

Findable: The GRAFT instrument should be easy to navigate, making it easy for participants.

Usable: The GRAFT instrument is easy to use in the post-transplant in-patient setting, easy to understand, and places minimal burden on clinicians with a short learning curve.

Two independent researchers coded each transcript and compared their respective codes, discussed, and organized them into a conceptual model. NVivo 12 provided the platform for coding the transcripts and conducting our qualitative analysis17.

To assess the reliability of the physicians' ratings using the GRAFT instrument, we assessed agreement in their classification of the same 10 heart transplant cases using Cohen's weighted

Kappa18. We used descriptive statistics to summarize the data collected through the SUS. STATA 15 provided the platform for our quantitative statistical analysis. Our threshold for statistical significance was a p-value of 0.05.

RESULTS

GRAFT Instrument

We developed a six-point ordinal classification instrument that classifies patients into normal, impaired but likely to achieve normal function, impaired and unlikely to achieve normal function, severely impaired but likely to achieve some function, severely impaired and unlikely to achieve some function, and irreversible graft failure (figure 1). We defined normal graft function as the same level of function one would expect in healthy native organs, impaired graft function as a level of function between normal and severe graft, and severely impaired graft function as a level of function associated with an appreciable reduction in graft longevity (figure 1). Through review of the United Nations for Organ Sharing's (UNOS) registry and the International Society of Heart and Lung Transplantation's (ISHLT) registry[19] and iterative consultations with clinical experts, we defined an appreciable reduction in graft longevity as a 25% probability of graft loss (recipient mortality or re-transplantation) at 1-year post transplantation. We defined irreversible graft failure as death due to graft dysfunction or assessment for re-transplantation.

We determined that an optimal time for the application of GRAFT and classification of patients is the time of discharge from index admission, 1-month for cases of prolonged hospital stay, or time of death for those who died prior to discharge within the first month post-transplant. The experts deemed these times to represent the period post-transplant where graft function is sufficiently stable for prognostication.

For individuals assessed and classified at the time of death, GRAFT-cardiac further classifies mortality as: due to graft dysfunction (e.g cases of irreversible graft failure), related to graft dysfunction (e.g. causes precipitated but not directly due to graft dysfunction), or un-related to graft dysfunction (e.g. strokes or infections unrelated to the graft).

The results of the systematic review informing our instrument will not be discussed here as the work was not part of this thesis. Our clinical experts determined and agreed that none of the known donor, transplant, and recipient characteristics modify the association between measures of graft function and the risk of 1-year mortality. The experts only inquired about the association between left ventricular ejection fraction (LVEF) or right atrial pressure (RAP) and mortality. The quality of the evidence for LVEF and RAP was deemed poor due to uncertainty about their timing of measurement. Therefore, none of the prognostic factors identified from our previous systematic review and meta-analysis informed our instruments thresholds for classification of patients. In response to the poor evidence identified, we conducted a primary retrospective cohort study (not part of this thesis) in which we assessed the association between serial measures of LVEF and RAP and mortality post-heart transplant using cox-regression modelling[9]. With this evidence we determined that an LVEF $\leq$45% or RAP >15 mmHg is associated with a 25% probability of graft-loss at 1-year post transplant. Based on recommendations from the American Society of Echocardiography, we set the threshold for normal LVEF at 62 $\pm$ 5%.

For individuals classified as impaired or severely impaired, through consultation with clinical experts, we distinguished between likely or unlikely to improve sub-categories based on the cause of graft dysfunction. Specifically, known and treatable causes such as (but not limited to) rejection or volume overload classify patients as likely to gain normal or any function for impaired

and severely impaired, respectively. Unknown or untreatable causes such as (but not limited to) primary graft dysfunction or cardiac allograft vasculopathy classify patients as unlikely to gain normal or any function for impaired and severely impaired, respectively.

User Testing

We interviewed five staff cardiologists from Toronto General Hospital. The physicians' years of experience with heart failure and heart transplantation ranged from 4 to 20 years.

Across these 5 physician participants who used GRAFT for classification of the 10 characteristic heart transplant cases, we observed high agreement in classifications of graft function/dysfunction with a Kappa of 0.87 (95% CI 0.62 – 1.00). The median score on the SUS 75 (range of 72.5 – 92.5) suggested acceptable usability.

Credibility

With regards to the credibility of the information and thresholds used for classification of the patients, two participants expressed concerns with regards to use of right atrial pressure (RAP). Participants felt that RAP measurements are dependent on renal function and as such elevated pressures may not be prognostically relevant:

"*I know there is a lot of data around RAP but from a patient to patient perspective I know overall, I'd like the RAP to be lower than higher but the measurement is so contextual. Hemodynamics are so contextual and it is hard to hang your hat on it. The RA pressure needs to be normalized to renal function or diuretic use or something.*"

"Hard to define right ventricular (RV) failure. This is something that hasn't been well established. I'm wondering if RAP alone is good enough to make a definition. Maybe there are other things

like Tricuspid Annular Plane Systolic Excursion (TAPSE) that you should consider to identify RV failure. I don't think, LVEF 40-60% has the same power or prediction as RAP of 9 mmHg. Most patients leave the operating room or intensive care unit with elevated RAP but doesn't mean there is impaired graft function."

Usefulness

Respondents described GRAFT as a tool that is useful for research but not for classification of patients clinical practice. Participants felt that GRAFT does not add any value to their clinical judgement, particularly due to lack of appropriate management strategies for each classification category:

"I think it is [useful], specially from a research perspective."
"what am I going to do? Like I classify someone as dead. What will I do with that?
I might label them but until we have therapy… categorization is useful for research. But until there is therapy for a category, then it doesn't make a difference. From research perspective this is very helpful because it might help with how you manage deceased donors."

"I'm not sure what this tool adds to my clinical judgment so I'm going to put It in the middle.
I'm automatically doing this. I wouldn't be using a tool to do what I already do.
It might be useful for training. Training fellows. For seasoned veterans, maybe, maybe not."

Accessibility

Our participants represented a wide range of experience in heart failure and heart transplantation. Irrespective of this variability, none shared any concern in their ability to project graft function recovery.

Findability

Our participants felt that the GRAFT instrument is sufficiently easy to navigate and use for classification of heart transplant recipients. Participants provided no feedback or suggestions for improvement with regards to findability.

Usability

Only one participant made comments specifically regarding the usability of our GRAFT instrument. The comments allowed us to make necessary revisions captured in the current version of the instrument. The participant felt that there were too many words presented on the instrument that limited their optimal usability.

"When I look at this tool is that there is a lot of words. It's a little bit busy. If I'm thinking about applying a tool, I want to be simplifying one way or another just for ease of access."

The four other participants did not experience and express any limitation regarding the usability of the GRAFT instrument.

DISCUSSION

Principle findings

We developed a 6-point generic instrument for classification of post-transplant graft function. Our instrument is intended for use at the time of discharge, at one month for patients still in hospital at that time, or time of death. The generic nature of this instrument serves the need of an outcome measurement tool for future randomized controlled trials addressing optimal deceased donor management strategies. To this end, the classification and definition for normal, impaired, severely impaired, and irreversible graft failure will be the same for all other organs (lungs, kidney, liver, and pancreases), with the exception of the specific thresholds. Thus far, for the cardiac version of the GRAFT instrument, we determined the specific evidence-based thresholds for each category, observed acceptable reliability, and showed acceptable usability amongst a sample of the intended users and will do so for other organ-specific GRAFT instruments.

Strengths and Limitations

GRAFT is a unique instrument that is sufficiently generic to capture a wholistic impact of any intervention applied to deceased donors. Through our close collaboration with clinical experts (heart, lung, liver, and kidney transplant experts from University Health Network and St. Joe's Hospital) we strived face validity for our instrument's measurement domains. The linkage of classification with evidence on patients' prognosis ensures criterion validity for our instrument. Our instrument's reliance on LVEF and RAP, two readily available markers of cardiac function[20], allow for easy and rapid classification of graft function/dysfunction. Furthermore, our user testing qualitative methods captured the positive experience of GRAFT's intended future users. This approach ensures that our cardiac version of the GRAFT instrument may be accepted and

well adopted by post-transplant cardiologist involved as outcome adjudicators for future trials on deceased donor management.

One limitation of our study is the small sample of users. This is best reflected by the wide 95% CI for our measure of agreement. If the lower bound of the 95% CI is true, our instrument's reliability is only just better than chance alone. When possible, the feedback received from the user testing allowed us to make further improvements to the GRAFT instrument. Although we believe that our modifications addressed concerns and suggestions raised by the participant, we are limited by uncertainty due to lack the limited number of participants in the user testing.

Within the user testing phase, we observed no concerns regarding participants' projection of graft recovery (accessibility). This may be partly due to extensive experience of staff cardiologists sampled from a high-volume transplant center. We are uncertain whether heart transplant fellows in training, or those from volume centres will share the same experience with regards to accessibility.

Regarding credibility, the participants raised concern regarding the use of RAP for classification of patients. Their major concern regarding its use was that in cases of renal dysfunction post-transplant, RAP may be an unreliable marker of cardiac dysfunction. RAP, however, is a readily available hemodynamic measure at the time of discharge, 1-month, or anytime during the index admission. The association between RAP and 1-year mortality informing our instrument was not adjusted for renal function of heart transplant recipients[9]. Previous studies that adjusted for renal function, however, did observe an independent association between RAP and mortality in a broad spectrum of patients with cardiovascular disease[21]. For future iterations of our

instrument we will strive to determine separate thresholds of RAP for those with and without renal dysfunction post-transplant.

In this study, we conducted our user testing with retrospective cases of heart transplant recipients. As such, some valuable information, available at the point of care to transplant physicians (such as donor right heart catheterization or echocardiography and donor medications), were not readily available to us through retrospective chart review. The lack of this information resulted in uncertainty about the likelihood of graft recovery, for impaired and severely impaired graft function (GRAFT categories 2 and 3). By studying the usability of GRAFT at the point of care, as opposed to using retrospective cases, may overcome this limitation and further gauge user's feedback on our guidance for subclassification of impaired or severely impaired graft function.

Comparison to other studies and implications

There are very few instruments developed for classification of graft function/dysfunction across various organ groups. For heart transplantation, in 2014 the ISHLT developed the primary graft dysfunction (PGD) instrument which is only limited to the first 24 hours post-transplantation[22]. Although various studies demonstrated validity of the PGD instrument[23 24], its use is limited to heart transplant recipients. Similarly, the lung transplant community utilizes their own PGD instrument[25], liver transplant community uses primary graft failure[26], and kidney transplant community utilizes delayed graft function[27]. All such instruments utilize their own unique definitions for classification categories. Such variability in their definitions creates difficulty in comparing, contrasting, and understanding the impact of any intervention on all organs procured across deceased donors. By creating a uniform definition, across all organ

groups, for each classification category of GRAFT, we've made comparisons across organ groups, and ultimately analysis of the impact of interventions, possible.

Previous instruments, such as PGD for heart transplantation, extensively rely on therapies given to patients for classification of patients[22]. This may potentially limit future endeavours in identifying therapies for graft recover. In contrast, in the cardiac version of GRAFT, we refrained from using therapy as a criterion for classification of patients. As such, our instrument may be further useful for planning future trials on optimal therapy for graft recovery post-transplant. This would, in turn, resolve the limitation our participants identified regarding the usefulness of GRAFT for clinical practice.

CONCLUSION

GRAFT is a 6-point generic rating of allograft function post-transplant. We designed GRAFT specifically as an outcome measure to be applied by transplant physicians in future randomized controlled trials of interventions for multiorgan donors. In this study, we developed and evaluated the cardiac version of GRAFT, demonstrating validity, reliability, and usability.

ACKNOWELDGEMENTS / FUNDING

None

DISCLOSURES

None of the authors have any conflict of interest to declare

REFERENECS

1. Bos EM, Leuvenink HG, van Goor H, et al. Kidney grafts from brain dead donors: Inferior quality or opportunity for improvement? *Kidney Int* 2007;72(7):797-805. doi: 10.1038/sj.ki.5002400 [published Online First: 2007/07/27]
2. Floerchinger B, Oberhuber R, Tullius SG. Effects of brain death on organ quality and transplant outcome. *Transplant Rev (Orlando)* 2012;26(2):54-9. doi: 10.1016/j.trre.2011.10.001 [published Online First: 2012/03/31]
3. Greer DM, Valenza F, Citerio G. Improving donor management and transplantation success: more research is needed. *Intensive Care Med* 2015;41(3):537-40. doi: 10.1007/s00134-015-3661-0 [published Online First: 2015/01/27]
4. Foroutan F, Alba AC, Stein M, et al. Validation of the International Society for Heart and Lung Transplantation primary graft dysfunction instrument in heart transplantation. *J Heart Lung Transplant* 2018 doi: 10.1016/j.healun.2018.12.007 [published Online First: 2019/01/16]
5. Foroutan F, Friesen EL, Clark KE, et al. Risk Factors for 1-Year Graft Loss After Kidney Transplantation: Systematic Review and Meta-Analysis. *Clin J Am Soc Nephrol* 2019 doi: 10.2215/CJN.05560519 [published Online First: 2019/09/22]
6. Foroutan F, Guyatt G, Friesen E, et al. Predictors of 1-year mortality in adult lung transplant recipients: a systematic review and meta-analysis. *Syst Rev* 2019;8(1):131. doi: 10.1186/s13643-019-1049-x [published Online First: 2019/06/05]
7. Foroutan F, Iorio A, Thabane L, et al. Calculation of absolute risk for important outcomes in patients with and without a prognostic factor of interest. *J Clin Epidemiol* 2019;117:46-51. doi: 10.1016/j.jclinepi.2019.08.012 [published Online First: 2019/10/08]
8. Foroutan F, Alba AC, Guyatt G, et al. Predictors of 1-year mortality in heart transplant recipients: a systematic review and meta-analysis. *Heart* 2018;104(2):151-60. doi: 10.1136/heartjnl-2017-311435 [published Online First: 2017/09/01]
9. Foroutan F, Ross HJ, Malik A, et al. Association Between Common Measures of Graft Function and Mortality Post Heart Transplant (HT). *The Journal of Heart and Lung Transplantation* 2018;37(4):S35. doi: 10.1016/j.healun.2018.01.066
10. Ericsson KA, Simon HA. Verbal reports as data. *Psychological Review* 1980;87(3):215-51. doi: 10.1037/0033-295X.87.3.215
11. Rosenbaum SE, Glenton C, Nylund HK, et al. User testing and stakeholder feedback contributed to the development of understandable and useful Summary of Findings tables for Cochrane reviews. *J Clin Epidemiol* 2010;63(6):607-19. doi: 10.1016/j.jclinepi.2009.12.013 [published Online First: 2010/05/04]
12. Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *J Clin Epidemiol* 2010;63(6):620-6. doi: 10.1016/j.jclinepi.2009.12.014
13. Willis GB, DeMaio T, Harris-Kojetin B. Is the Bandwagon Headed to the Methodological Promised Land? Evaluating the Validity of Cognitive Interviewing Techniques. Sirken MG, Herrmann D, Schechter S, Schwarz N, Tanur J, Tourangeau R, editors Cognition in Survey Research. New York: John Wiley 1999:133-54.
14. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Lyall TI, eds. Usability Evaluation in Industry: Taylor and Francis 1996:189 - 94.

15. Elo S, Kyngas H. The qualitative content analysis process. *Journal of advanced nursing* 2008;62(1):107-15. doi: 10.1111/j.1365-2648.2007.04569.x
16. Morville P. User Experience Design 2004 [Available from: http://semanticstudios.com/user_experience_design/2019.
17. Ltd QIP. NVivo qualitative data analysis software, 2018.
18. KAPPAETC: Stata module to evaluate interrater agreement [program]. S458283 version: Boston College Department of Economics, 2016.
19. Khush KK, Cherikh WS, Chambers DC, et al. The International Thoracic Organ Transplant Registry of the International Society for Heart and Lung Transplantation: Thirty-sixth adult heart transplantation report - 2019; focus theme: Donor and recipient size match. *J Heart Lung Transplant* 2019;38(10):1056-66. doi: 10.1016/j.healun.2019.08.004 [published Online First: 2019/09/25]
20. Costanzo MR, Dipchand A, Starling R, et al. The International Society of Heart and Lung Transplantation Guidelines for the care of heart transplant recipients. *J Heart Lung Transplant* 2010;29(8):914-56. doi: 10.1016/j.healun.2010.05.034 [published Online First: 2010/07/21]
21. Damman K, van Deursen VM, Navis G, et al. Increased central venous pressure is associated with impaired renal function and mortality in a broad spectrum of patients with cardiovascular disease. *J Am Coll Cardiol* 2009;53(7):582-88. doi: 10.1016/j.jacc.2008.08.080 [published Online First: 2009/02/14]
22. Kobashigawa J, Zuckermann A, Macdonald P, et al. Report from a consensus conference on primary graft dysfunction after cardiac transplantation. *J Heart Lung Transplant* 2014;33(4):327-40. doi: 10.1016/j.healun.2014.02.027 [published Online First: 2014/03/26]
23. Foroutan F, Alba AC, Stein M, et al. Validation of the International Society for Heart and Lung Transplantation primary graft dysfunction instrument in heart transplantation. *J Heart Lung Transplant* 2019;38(3):260-66. doi: 10.1016/j.healun.2018.12.007 [published Online First: 2019/01/16]
24. Foroutan F, Ross HJ. Primary Graft Dysfunction: The Devil Is in the Details. *Transplantation* 2019;103(2):229-30. doi: 10.1097/TP.0000000000002221 [published Online First: 2018/05/15]
25. Christie JD, Carby M, Bag R, et al. Report of the ISHLT Working Group on Primary Lung Graft Dysfunction part II: definition. A consensus statement of the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant* 2005;24(10):1454-9. doi: 10.1016/j.healun.2004.11.049 [published Online First: 2005/10/08]
26. Chen XB, Xu MQ. Primary graft dysfunction after liver transplantation. *Hepatobiliary Pancreat Dis Int* 2014;13(2):125-37. [published Online First: 2014/04/02]
27. Siedlecki A, Irish W, Brennan DC. Delayed graft function in the kidney transplant. *Am J Transplant* 2011;11(11):2279-96. doi: 10.1111/j.1600-6143.2011.03754.x [published Online First: 2011/09/21]

Figure 1 – Development process

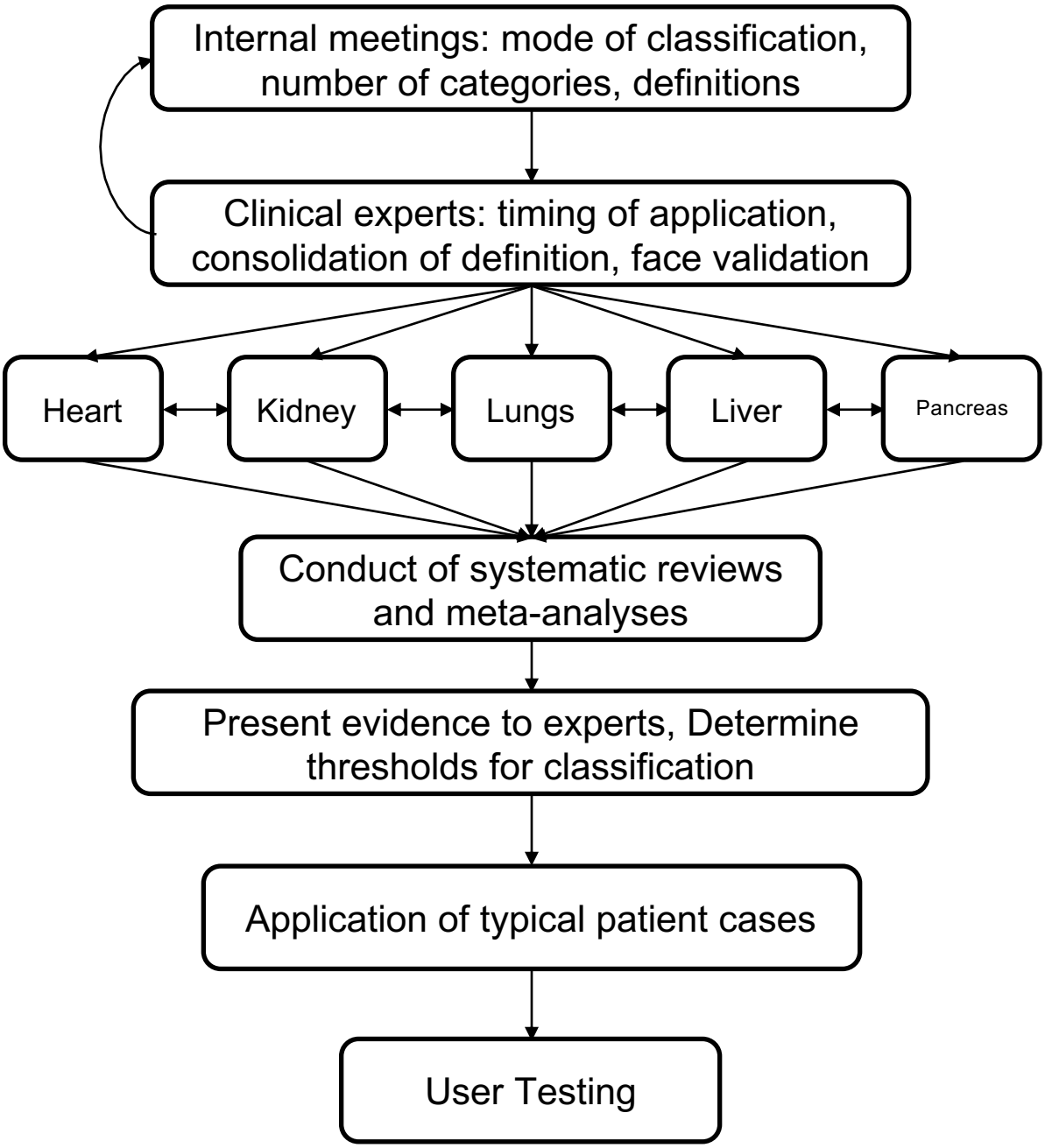


Figure 2 – Cardiac version of the GRAFT instrument

Cardiac Version of GRAFT (Generic Rating of Allograft Function post-Transplant)

Patient Info

Patient ID	**Name (Last, First)**	**Date of Transplant**
_____	_____	_____

Timing of Application (please select one)

On day of discharge 1-month post-transplant At the time of death

GRAFT Category & Definition	GRAFT Subcategory	Guide & Thresholds
1. Normal graft function	→ → → → → → **[No subcategory]**	LVEF >60 ± 5% c RAP <8 mmH
2. Impaired graft function *Level of function between normal and severe graft dysfunction*	**A.** **Likely to achieve normal function**	LVEF 40 – 60% c RAP 8 – 15 mmH Cause: rejection volume overload known and treatable aetiology
	B. **Unlikely to achieve normal function**	LVEF 46 – 60% c RAP 8 – 15 mmH Cause: Primary Graft Dysfunction Unknown aetiology Known but un-treatable aetiology
3. Severely impaired graft function *An appreciable reduction in graft longevity: **25%** probability of graft loss at **1-year**.*	**A.** **Likely to gain some function**	LVEF ≤ 45% c RAP > 15 mmH Cause: rejection volume overload known and treatable aetiology
	B. **Unlikely to regain any function**	LVEF ≤ 45% c RAP > 15 mmH Cause: Primary Graft Dysfunction Unknown aetiology Known but un-treatable aetiology
4. Irreversible graft failure	→ → → → → → **[No subcategory]**	Mortality due to graft failure / Re-transplantation

Cause of Death (please select one)

Due to graft dysfunction Related to graft dysfunction Unrelated to graft dysfunction

APPENDIX A – Interview Guide

Think aloud (~30 minutes)

During this phase, participants will engage in the following task:

I have just explained to you how to use the GRAFT instrument to classify patients according their organ function post-transplant. Now, do you have any questions about the instrument or the process in rating patients?

Thanks for your questions.

Now, please read the patient scenario provided. As you are reading through the scenario, you are likely to have thoughts about how the information will lead you to use the GRAFT instrument. You may have thoughts about how the description is unclear, or is missing important information that you would like in applying the GRAFT instrument. Please think aloud, and tell me the thoughts you are having about the scenario, and about how you are thinking about how the scenario will apply to the addressing the GRAFT instrument.

After reading all the information provided on the clinical case, please use the GRAFT instrument to classify the graft function depicted in the scenario. Please continue to think aloud as you ponder how to classify the patient on the GRAFT instrument after you complete reading the scenario.

Prompts for think aloud phase:

We will ask the participant to elaborate on his/her comments when appropriate or when they fall silent for a few seconds.

Prompt 1: What are you thinking? Does the scenario have the information you need? What are you thinking now about the GRAFT rating you will be making?

Prompt 2: Can you describe what it is in the scenario that you are thinking about now?

Semi-structured open-ended questions (~30 minutes)

After completion of the think aloud phase, we will ask the participants the following set of open-ended questions:

1. Credible: How did you feel about the scenario? Did it give you all the information you need to make your GRAFT rating? Could the information have been presented in a better way?

   What about the GRAFT instrument itself? Do you think it's a reasonable approach to rating heart function after transplant?
2. Useful: Is there any other way the patient scenario could be improved to allow you to make the GRAFT rating? What about the GRAFT instrument? Is there any way it could be improved?
3. Accessible: Please tell us challenges you faced when making your GRAFT rating.
4. Findable: Can you please express your additional thoughts and opinions on using the GRAFT instrument, specifically with regards to how easy it is to understand and use?

Chapter 9: Discussion and Concluding Remarks

In this thesis I presented the development of GRAFT for heart transplantation. Although our focus was the cardiac version of GRAFT, similar methods are being implemented for the development lung, renal, hepatic, and pancreatic version of GRAFT as well. This thesis presents the obstacles and solutions when developing this generic measurement instrument.

Function of the transplanted organs is very specific. To date, each organ group developed instrument for classification of graft function focused exclusively on their own organ. For example, The International Society of Heart and Lung Transplantation's (ISHLT) Primary Graft Dysfunction (PGD) instrument defines and classifies the severity of a well-recognized post-transplant occurrence[1]. Prior to the publication of ISHLT's PGD instrument, variation in the classification of PGD created difficulty in understanding the possible underlying causes[2]. With a valid consensus instrument, we can better study the risk factors, histology, and biological manifestation of PGD. Although we demonstrated the validity of ISHLT's PGD instrument (chapter 2), its use in the setting of randomized controlled trials on management strategies of deceased donors.

Current approaches make it impossible to summarize the impact of an intervention across organs using a single measure. Sample size will inevitably be limited in randomized trials of interventions for deceased organ donors and an approach that maximizes the power of an analysis may be crucial. GRAFT, which will allow aggregation of results across organs, is likely to play an important role in achieving optimal analytic power.

GRAFT's strategy for achieving it goal is by linking classification of patients to prognosis. Our definition for severely impaired graft function represents patients with an appreciable reduction in graft longevity (25% probability of graft loss at 1-year post-transplant). This uniform definition across all organ groups overcomes barriers in comparing and contrasting classification of graft function across different organ groups. The goal of, in future randomized controlled trials of deceased donor management, enhancing understanding of how a single intervention variably impacts hearts, lungs, kidneys, liver, and pancreas will require completion of GRAFT for all other organ groups.

In a previous systematic review and meta-analysis of prognostic factors for 1-year mortality post hear-transplantation, we summarized all donor, transplant, and recipient factors that impact graft longevity[3]. For the development of the heart version of GRAFT first edition, none of these characteristics were deemed necessary for classification of patients. This judgement, however, may not be applicable to development of GRAFT for lungs, kidney, liver, and pancreas. To this end, we needed evidence, in the form of systematic review and meta-analysis, on prognostic factors for the other organ groups as well. Within this thesis, chapters 2-4 demonstrated the methods necessary for conduct of such a systematic review. We demonstrated the methods for lungs and kidneys and identified a number of prognostic factors that influence graft longevity. At the moment, the results of these reviews are informing a panel of clinical experts identifying thresholds that will be used for classification of patients in each of GRAFT's categories.

As our clinical experts review the evidence from systematic reviews and meta-analyses of prognostic factors, how certain we are on the effect of each prognostic factor on graft longevity

will influence their decision for using that prognostic factor to inform the threshold for classification of patients. Beyond the development of our instrument, users of evidence on prognostic factors may either be interested in their use for study planning and analysis or for clinical decision-making. Prior to this thesis, Grading of Recommendations Assessment, Development and Evaluation (GRADE) had only provided guidance for assessing certainty of evidence on overall prognosis of particular group of patients[4]. In chapter 5 we demonstrated that the same principles GRADE proposed for body of evidence addressing treatment and overall prognosis work well in assessing bodies of evidence regarding individual prognostic factors.

Through the conduct of this thesis, we observed that, by and large, authors of primary studies on prognostic factors report only the relative effect of the factor. Although relative effects are useful for understanding prognostic factors across a wide range of patients, their use in helping patients plan for their care/future is limited. For clinical decision making, relative effects must be converted to absolute risks. Calculation of absolute risk, however, is challenging when we don't know the risk of the outcome of interest in those without the prognostic factor.

The calculator we developed in chapter 6 provides a framework in which users of medical literature can derive absolute risks in those with and without a prognostic factor of interest, using the relative effect and prevalence of the factor, along with the overall risk of the outcome. Such calculations contextualize the relative effect estimates for prognostic factors, allowing for a more informed decision-making process. Authors of future systematic reviews would well serve their clinician readers by conducting the proposed calculations, or at least summarizing and reporting all the three components needed for the calculation, so that users of medical literature may use our framework to calculate the relevant absolute risks at the decision-making point.

In the end, using all the methods and techniques developed in this thesis, we completed the cardiac version of GRAFT. For the cardiac version of GRAFT, we demonstrated acceptable validity, reliability, and usability through the formal mixed methods user testing study (chapter 7). If GRAFT is widely adopted as an outcome measure for all future trials in deceased donor management, it will enhance the feasibility of trials in which interventions directed at deceased donors are tested across organ systems.

References

1. Kobashigawa J, Zuckermann A, Macdonald P, et al. Report from a consensus conference on primary graft dysfunction after cardiac transplantation. *J Heart Lung Transplant* 2014;33(4):327-40. doi: 10.1016/j.healun.2014.02.027 [published Online First: 2014/03/26]
2. Segovia J, Cosio MD, Barcelo JM, et al. RADIAL: a novel primary graft failure risk score in heart transplantation. *J Heart Lung Transplant* 2011;30(6):644-51. doi: 10.1016/j.healun.2011.01.721 [published Online First: 2011/04/08]
3. Foroutan F, Alba AC, Guyatt G, et al. Predictors of 1-year mortality in heart transplant recipients: a systematic review and meta-analysis. *Heart* 2018;104(2):151-60. doi: 10.1136/heartjnl-2017-311435 [published Online First: 2017/09/01]
4. Iorio A, Spencer FA, Falavigna M, et al. Use of GRADE for assessment of evidence about prognosis: rating confidence in estimates of event rates in broad categories of patients. *BMJ* 2015;350:h870. doi: 10.1136/bmj.h870 [published Online First: 2015/03/18]