Preventable errors in the performance and reporting of echocardiography and right heart catheterization can lead to diagnosis errors for pulmonary hypertension with significant implications for patients

by

Alexandra Saunders

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Translational Medicine

Department of Medicine

University of Alberta

Abstract

Pulmonary hypertension (PH) is common and PH subgroups have vastly different mortality and treatment, particularly pulmonary arterial hypertension (PAH) versus PH secondary to heart failure with preserved ejection fraction (HFpEF). Both transthoracic echocardiography (TTE) and right heart catheterization (RHC) are needed, as diagnosis requires mean pulmonary artery pressure (mPAP) >20 mmHg either with pulmonary artery wedge pressure (PAWP) <15 mmHg for PAH or >15 mmHg for HFpEF. PAWP and mPAP can only be measured by RHC. However, TTE is what first identifies PH, triggering referral to a specialized PH center. We hypothesized that human errors in the performance and reporting of TTEs and RHC are prevalent, potentially leading to misdiagnosis of PH and its subgroups.

We re-analyzed TTEs and RHCs of 263 consecutive new patients referred to our PH program during a 5-year period. We also compared the inferred diagnosis from the referring TTE and the subsequent RHC reports to the diagnosis made after correcting for errors found. We identified numerous preventable errors in the performance and reporting of both tests. There was a poor correlation between the parameters measured by both tests (e.g., systolic PAP). The referral TTE reports missed or overcalled PH in 44 patients. The RHC, mostly by mistakes of the PAWP, led to misdiagnosis in 41 patients, (21 with true PAH labelled as HFpEF and 20 with true HFpEF labelled as PAH).

TTE errors may delay referrals and RHC errors may lead to misdiagnosis and applying wrong therapies to the wrong groups. As PAH therapies are extremely expensive, this also impacts the health care system. Primary care physicians need to be on alert for such errors and referral centers need to promote quality improvement programs that could eliminate these errors.

Preface

This thesis is an original work by Alexandra Saunders. No part of this thesis has been previously published, though publication will be eventually pursued. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board on January 22, 2022.

Acknowledgements

True mastery of a topic is never finished, but I want to take a moment to acknowledge all the people in my life who have helped me get to this level in my training and now transitioning into my role as staff, and to thank them in advance for all the help that is to come.

To my mentor, Dr Evangelos Michelakis: he was one of the reasons that I chose to train at the University of Alberta for my cardiology residency, as I wanted to have the opportunity to work and learn from a world expert in pulmonary hypertension. Suffice to say that he has taught me a lot more than just pulmonary hypertension physiology; he has helped me to also think critically through problems beyond the realm of PH. The Translational Medicine Masters course was one part of this, but our weekly chats and discussions also helped build this skill. This critical lens will not only help me in patient care and research, but with other large life decisions. I look forward to continuing to work with him beyond this time of my "PH apprenticeship".

To my parents, of which I am lucky enough to have four: for always believing in my abilities, while also letting me discover for myself that any limits to my potential or dreams are often self-imposed – i.e. showing me the power of dreaming big. Hanging in my office is a little poem my mom wrote to me when I graduated high school, and it really summarizes the transformation of a not-so-tame spirit into a person with energy that is focused and driven. I look at it when I need reminding of myself, and her ongoing quiet support (that I know is still there, even though she cannot express it with anything other than a smile). Each of my parents have made a little part of me the way I am today.

To my friends: for accepting both who I am at heart, but also how I have changed along the way, and how much of an impact they have had on those changes. My handful of treasured friends are like the chapters of a book; my time is marked by the time before, and the time after they came into my life. I truly appreciate the time they have taken to be with me, especially as life moves along and I see how limited that time can be.

To my husband: for growing with me, when I needed him to, through this long journey, particularly with the surprises of extra years and long-distance training. For being excited for our next chapter. For having to listen to all my insecurities and doubts and never making me feel judged for them. I am also very excited for what this next chapter will bring and will certainly need a strong teammate at my side.

Table of Contents

Title Page i
Abstract ii
Preface iii
Acknowledgements iv
Table of Contents v
List of Tables vi
List of Figures viii
Glossary of Terms xv
Introduction 1
Methods 5
Results 8
Discussion 11
Clinical Perspectives 13
References 15

List of Tables

Table 1. Data collected for TTE and RHC studies and their rationale.

Parameter Reported on TTE	**Rationale**
University vs community site	For comparison of quality of studies.
RVSP (i.e. sPAP)	Typically, if >40 mmHg, PH is listed in the conclusion of the report.
Indication of whether TR velocity envelope is adequate (checking of envelope or if checked for in 3 views)	Raises suspicion for potentially unreliable estimates of sPAP.
RAP (based on IVC distensibility)	This is needed to calculate sPAP, as opposed to an arbitrary value of 5 - 15 mmHg used in some centers.
RV free wall thickness	Suggestive of chronic PH or likelihood of compensated RV.
RV size (RV dilation grade)	Suggestive of likelihood of decompensated RV.
RV contractility (TAPSE)	Indicates RV decompensation if <2.0 cm.
Parameters Reported on RHC	**Rationale**
Right atrial pressure (RAP)	For comparison to TTE RAP and to determine if zeroing was appropriate (as this value should not be <-3 mmHg).
sPAP, dPAP, mPAP	For comparison to TTE RVSP and to diagnose PH and to confirm quality of PAWP (as dPAP should be >PAWP).
PAWP waveform quality (O_2 saturation to confirm wedge position, average of 3 beats, and end-expiratory recording)	Validates the accuracy of the measurement and raises suspicion that the value of PAWP, and therefore PVR, may not be accurate.
LVEDP	Validates the accuracy of the reported PAWP.
CO (thermodilution, Fick)	The accuracy of CO varies between thermodilution and Fick methods, and thus so does the calculation of PVR.
PVR value and whether this is automatically entered by the catheterization lab algorithm or manually calculated	Used in the diagnosis of PH and vulnerable errors in both PAWP and CO.

Table 2. Frequency of potential measurement errors (left) and number of misses/misclassifications of PH because of these errors.

Parameters on TTE	**Number of patients (%)**	**Type of misdiagnosis from TTE errors** • PH missed: 9 (4) • PH overcalled: 35 (14) ---------- **Type of misdiagnosis from RHC errors** • True PAH labelled as HFpEF due to error in PAWP: 21 (8) • True HFpEF labelled as PAH: 20 (8)
Poor quality of TR envelope not reported	61 (34)	
TR not checked in 3 views	21 (12)	
RAP not appropriately based on IVC	23 (9)	
RV free wall thickness not reported	245 (97)	
RV size not reported	9 (4)	
RV contractility not reported (and without rationale, like poor image quality)	41 (16)	
Parameters on RHC		
PAWP position not confirmed using PAWP O_2 saturation	252 (100)	
PAWP not reported at end-expiratory stage (when respiratory variation was recorded)	60 (24)	
PAWP recorded during a breath hold	81 (32)	
PAWP not the average of 3 beats	26 (10)	
PAWP > 5 mmHg difference from LVEDP	48 (38)	
PAWP > dPAP	25 (10)	
CO method not done using thermodilution	217 (86)	

List of Figures

Figure Legends

Figure 1: All TTE RVSP values compared to patient-matched RHC sPAP values, with black dots marking studies performed at the university echo lab, and orange dots at outside centers echo labs. All RHC were done at the University site.
1A: Top: A Pearson correlation is shown. The values only correlated moderately and with the same r value of 0.6, whether they were University or community TTEs. **Bottom**: A Bland-Altman analysis of the paired TTE/RHC values shows a wide range of disagreement between the 2 tests, higher than what is considered clinically "acceptable". The red box shown patients where TTE would have missed the diagnosis of PH (typically reported if RVSP is >40 mmHg) since the same patients had higher values of sPAP in the RHC. The blue box represents patients where TTE stated that there was PH, when there was no PH in the RHC (sPAP<40 mmHg).
1B: The same Pearson correlation is shown in subgroups where the 2 tests were done within 2 days from each other, between 2 days and 2 weeks, and within >2 weeks.

Figure 2: Scatterplot of RHC-based calculated PAPi vs TTE TAPSE, showing no correlation **(A)**. There is also no correlation between PAPi and RV size at end-diastole reported on the TTE **(B)**.

Figure 3: Pearson correlations and Bland-Altman analyses for comparisons of the LVEDP against various forms of the PAWP measurements:
A: LVEDP versus PAWP reported in the same study of the same patient.
B: LVEDP versus the corrected PAWP during our re-analysis of the RHC recorded pressure tracings. Comparing A to B shows that the PAWP traces were not interpreted correctly, as our reanalysis increased the r value toward LVEDP (i.e. the gold standard for the accuracy of the PAWP in the absence of mitral stenosis, which was excluded in our studies)
C: LVEDP versus the reported PAWP when no respiratory variation was stated in the report and the PAWP was recorded during a breath hold.

D: LVEDP versus the reported PAWP when the report clearly included traces showing respiratory variation of the PAWP value. Comparing C to D shows that the physicians sensitized to correctly recording and reporting PAWP were much closer to the gold standard value (i.e., LVEDP).

Red boxes indicate patients where the PAWP was reported as high (suggesting for example HFpEF) while the LVEDP was <15 mmHg, thus missing a diagnosis of PAH. Blue boxes indicate patients where the PAWP was reported as low, suggesting PAH, while the true LVEDP was >15 mmHg, giving an incorrect label of PAH and missing PH secondary to HFpEF.

Figure 4. Pearson correlation and Bland-Altman analysis between Fick vs thermodilution performed in the same study of the same patient show poor correlation and wide degree of disagreement.

Figure 5: Pearson correlations of WHO functional class (**A**) and 6-minute walk distance (**B**) assessed close to the time of RHC to either indirect Fick-measured (**left**) or thermodilution-measured cardiac output (**right**). The Fick cardiac output did not correlate at all with these two clinical paraments that are expected to relate with and unfortunately was the method used in the majority of RHCs.

Figure 1

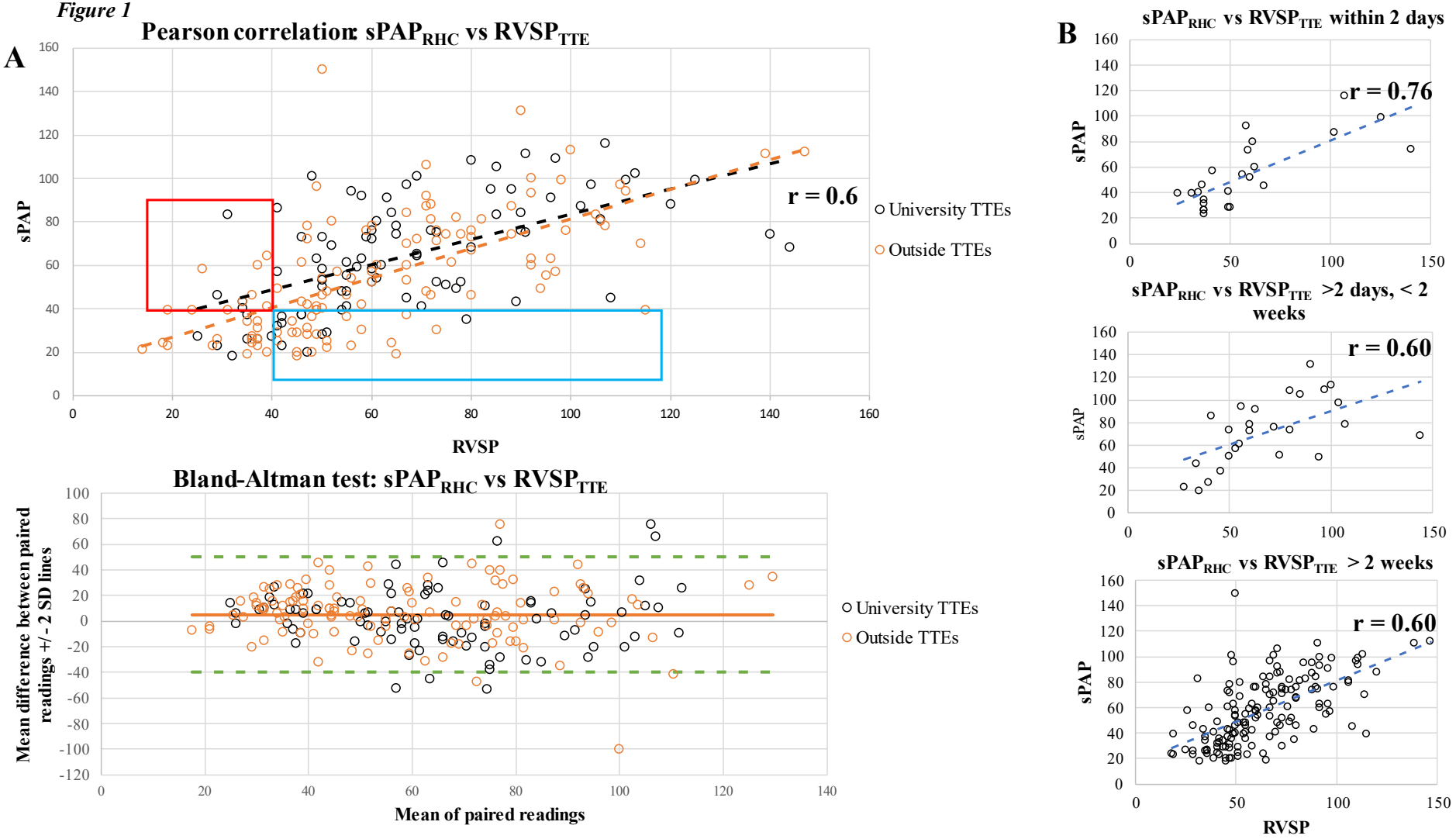

A
Pearson correlation: sPAP$_{RHC}$ vs RVSP$_{TTE}$
sPAP
RVSP
r = 0.6
University TTEs
Outside TTEs
Bland-Altman test: sPAP$_{RHC}$ vs RVSP$_{TTE}$
Mean difference between paired readings +/- 2 SD lines
Mean of paired readings
University TTEs
Outside TTEs
B
sPAP$_{RHC}$ vs RVSP$_{TTE}$ within 2 days
r = 0.76
sPAP$_{RHC}$ vs RVSP$_{TTE}$ >2 days, < 2 weeks
r = 0.60
sPAP$_{RHC}$ vs RVSP$_{TTE}$ > 2 weeks
r = 0.60
sPAP
RVSP

Figure 2

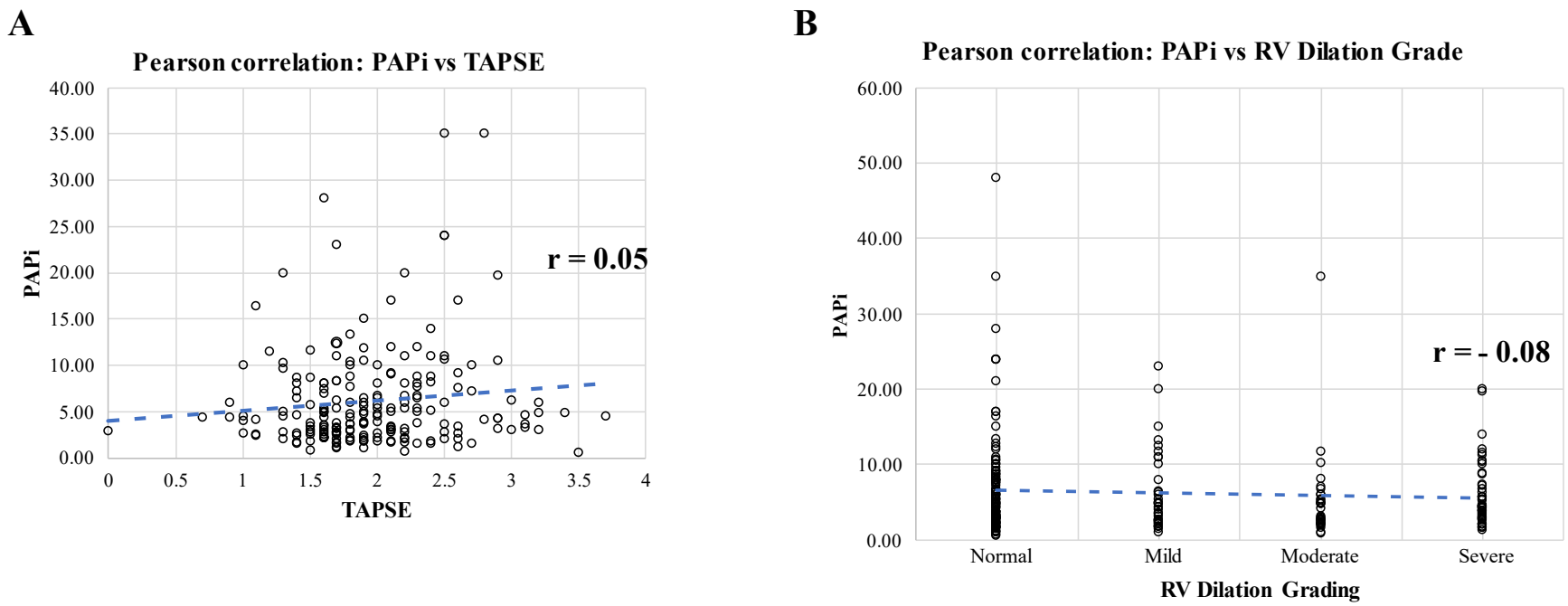

A
Pearson correlation: PAPi vs TAPSE
PAPi
TAPSE
r = 0.05
40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
0
0.5
1
1.5
2
2.5
3
3.5
4
B
Pearson correlation: PAPi vs RV Dilation Grade
PAPi
RV Dilation Grading
r = - 0.08
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Normal
Mild
Moderate
Severe

Figure 3

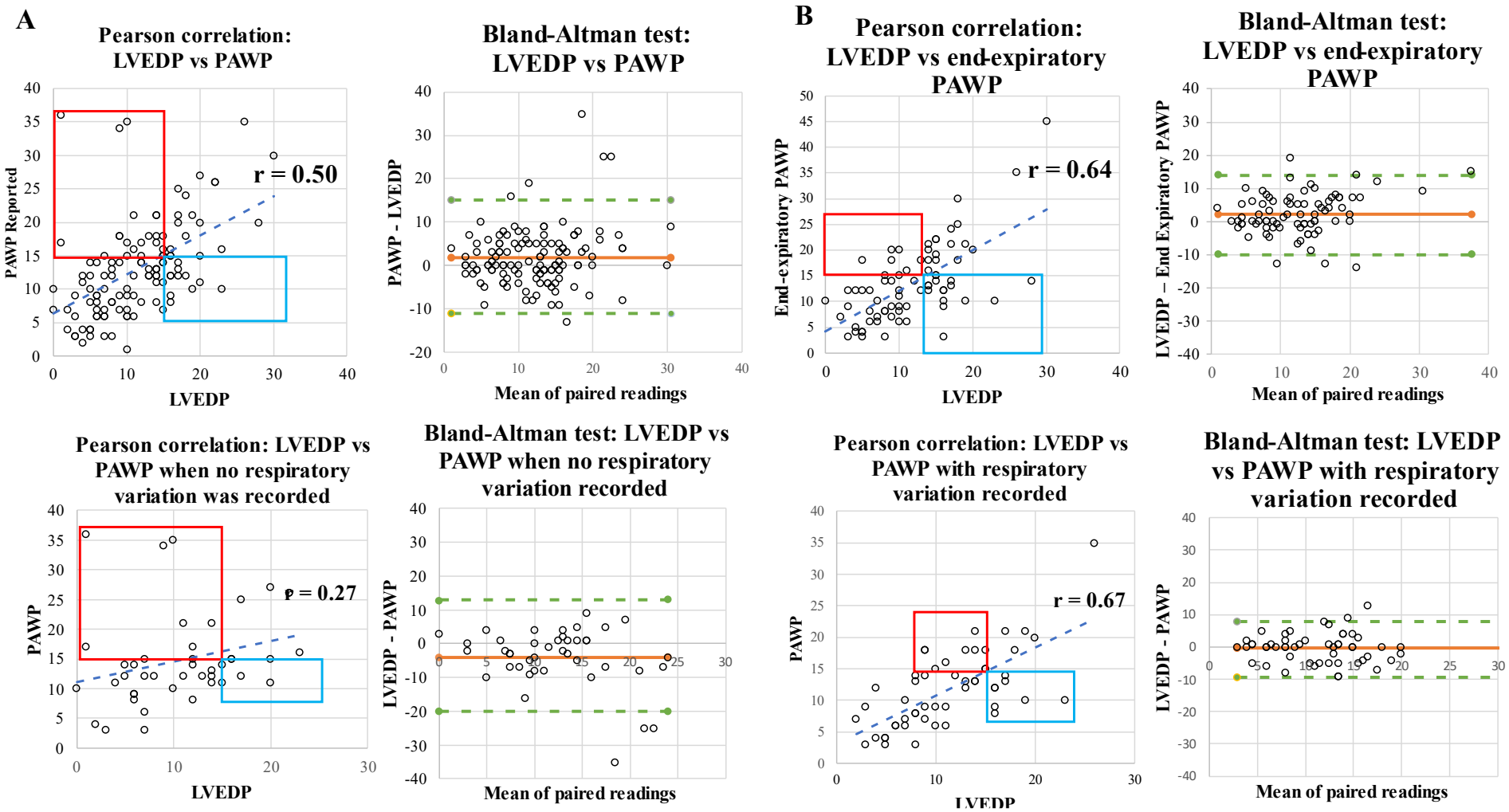
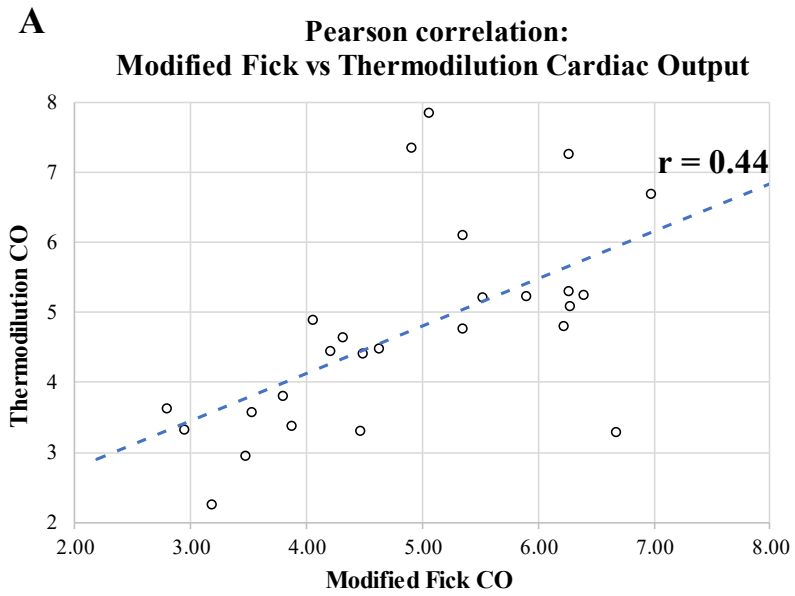

A
Pearson correlation: LVEDP vs PAWP
PAWP Reported
LVEDP
r = 0.50
Bland-Altman test: LVEDP vs PAWP
PAWP - LVEDP
Mean of paired readings
B
Pearson correlation: LVEDP vs end-expiratory PAWP
End-expiratory PAWP
LVEDP
r = 0.64
Bland-Altman test: LVEDP vs end-expiratory PAWP
LVEDP – End Expiratory PAWP
Mean of paired readings
Pearson correlation: LVEDP vs PAWP when no respiratory variation was recorded
PAWP
LVEDP
r = 0.27
Bland-Altman test: LVEDP vs PAWP when no respiratory variation recorded
LVEDP - PAWP
Mean of paired readings
Pearson correlation: LVEDP vs PAWP with respiratory variation recorded
PAWP
LVEDP
r = 0.67
Bland-Altman test: LVEDP vs PAWP with respiratory variation recorded
LVEDP - PAWP
Mean of paired readings

Figure 4


A
Pearson correlation:
Modified Fick vs Thermodilution Cardiac Output
r = 0.44
Thermodilution CO
Modified Fick CO
2.00
3.00
4.00
5.00
6.00
7.00
8.00
2
3
4
5
6
7
8

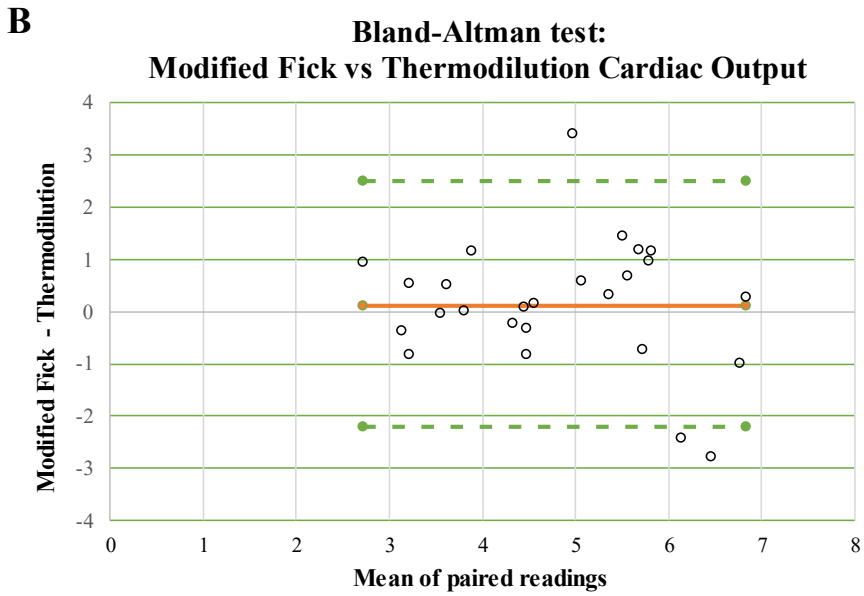
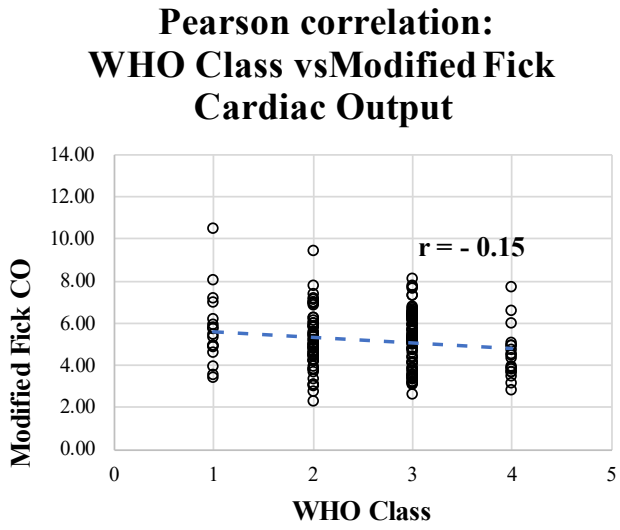
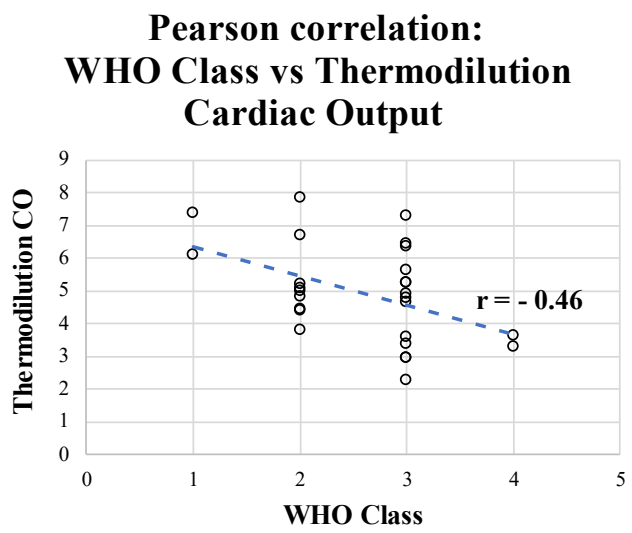

B
Bland-Altman test:
Modified Fick vs Thermodilution Cardiac Output
Modified Fick - Thermodilution
Mean of paired readings
0
1
2
3
4
5
6
7
8
4
3
2
1
0
-1
-2
-3
-4

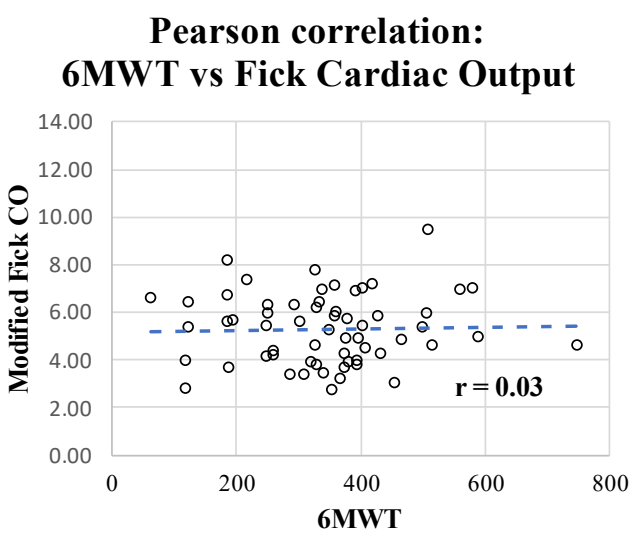
Figure 5

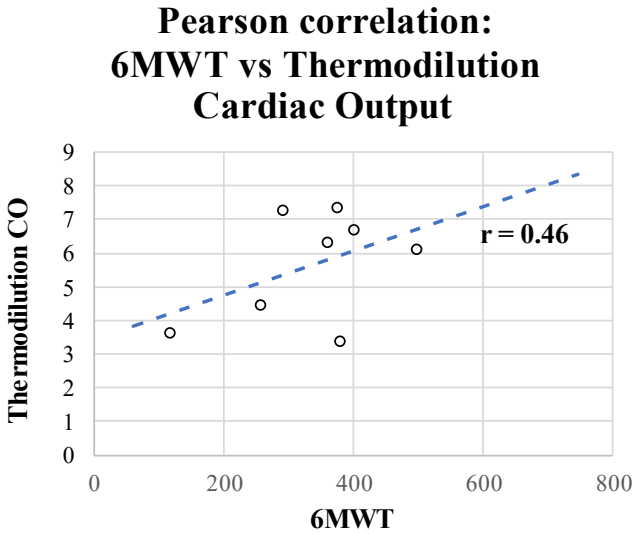
A


Pearson correlation:
WHO Class vsModified Fick Cardiac Output
Modified Fick CO
WHO Class
r = - 0.15
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0
1
2
3
4
5


Pearson correlation:
WHO Class vs Thermodilution Cardiac Output
Thermodilution CO
WHO Class
r = - 0.46
9
8
7
6
5
4
3
2
1
0
0
1
2
3
4
5

B


Pearson correlation:
6MWT vs Fick Cardiac Output
Modified Fick CO
6MWT
r = 0.03
14.00
12.00
10.00
8.00
6.00
4.00
2.00
0.00
0
200
400
600
800


Pearson correlation:
6MWT vs Thermodilution Cardiac Output
Thermodilution CO
6MWT
r = 0.46
9
8
7
6
5
4
3
2
1
0
0
200
400
600
800

Glossary of Terms

Key words: Pulmonary arterial hypertension, pulmonary hypertension, heart failure with preserved ejection fraction, quality improvement, right heart catheterization, echocardiography

Abbreviations List:

6MWT – six-minute walk test
HFpEF – heart failure with preserved ejection fraction
PAH – pulmonary arterial hypertension
PAP – Pulmonary artery pressure (with systolic(s), diastolic(d), and mean(m) modifiers)
PAWP – pulmonary artery wedge pressure
PH – pulmonary hypertension
RHC – right heart catheterization
RVSP – right ventricular systolic pressure
TTE – transthoracic echocardiogram
WHO-FC – World Health Organization – Functional Capacity

Introduction

Pulmonary hypertension (PH) is a disease with high morbidity and mortality, with the prognosis varying among its subgroups. Even with therapy, pulmonary arterial hypertension (PAH; Group 1) mortality is high with a 59% 5-year survival, worse than many metastatic cancers(1, 2). Medical therapy can help to move patients into a lower risk group, but the mortality even for this group is up to 16% over 3 years(3). Therapies are very expensive and often with significant side effects, but unable to reverse the disease or significantly prolong survival. Late diagnosis also worsens mortality by >20%(4).

PH is an umbrella term for a group of heterogeneous diseases that cause elevations in the right-sided pressures of the heart, with elevations in pulmonary artery (PA) pressures and the potential to eventually cause detrimental effects like RV overload, failure, and reduced cardiac output/cardiac index (CO/CI). Traditionally these diseases have been placed into five groups(5, 6). Group 1 is pulmonary arterial hypertension (PAH) and is when the pulmonary arteries undergo a state of proliferation, apoptosis resistance, and metabolic dysregulation(7) causing a dramatic increase in PA pressures related to the luminal narrowing and increase is resistance across the circuit. Treatment of PAH is currently targeted towards attempts to vasodilate these narrowed vessels through use of phosphodiesterase group 5 (PDE-5) inhibitors (like sildenafil), endothelin receptor antagonists (like bosentan), or synthetic prostacyclins (like with parenteral route via epoprostenol or an oral route via selexipag)(6). There are recently approved therapies(8) as well as ongoing research(9) in that may target the proliferative aspect of PAH, but the true clinical impact of these medications remain to be seen.

Group 2 is PH related to left heart disease and occurs when left ventricular filling pressures are increased and transmit these elevated pressures to the right side. It can be isolated post-capillary PH, but often these patients will also develop a degree of pulmonary vasculature remodeling resulting in a combined post- and pre-capillary phenotype(6). Increased left-sided filling pressures can be related to relatively asymptomatic diastolic dysfunction or to a clinical syndrome of heart failure, which can be with reduced or preserved ejection fraction(10). The management of reduced ejection has been well established with the four pillars of therapy, with angiotensin-converting enzyme inhibitors/angiotensin 2 receptor blockers/angiotensin receptor/neprilysin inhibitors, beta blockers, mineralocorticoid receptor antagonists, and diuretics as needed(10). Recently, promising classes of drugs (like SGL2 inhibitors (13)), have been shown to be effective in the treatment of

heart failure with preserved ejection fraction (HFpEF), with ongoing research into further therapies.

Group 3 PH is from lung diseases, like chronic obstructive pulmonary disease or interstitial lung disease. It is the second most common form of PH and is unfortunately as difficult to manage as the underlying lung disease associated with it – which are often irreversible. The mechanism of PH in this group is multifactorial and related to changes in the lung parenchyma as well as vascular abnormalities/remodeling (which is often affected to a greater extent than one would expect from the parenchymal disease)(11). Treatment is often just management of the underlying disease, but use of PDE-5 inhibitors can be considered(6).

Group 4 PH is from obstructions to the pulmonary artery, most commonly from chronic thromboembolic pulmonary hypertension, but can also be from malignant obstructions. Treatment of this is with a combination of surgery (thromboendarterectomy, if they are a good candidate based on anatomy, comorbidities, and clot burden), medication (riociguat or PDE-5 inhibitors), and in some cases balloon angioplasty(6).

Group 5 is a mixed group of diseases with multifactorial mechanisms that do not quite fit into the other groups, and includes conditions like sarcoidosis, glycogen storage diseases, and myeloproliferative disorders. Management of this group is based on the underlying condition, as well as some trials of PAH therapies as guided by their local expert centre(6).

Getting the diagnosis/PH subgroup correct is key to determining the appropriate treatment strategy; as outlined above, they can be quite different, and the different subgroups have varying morbidity and mortality. In group 2 PH, for example, the 5-year survival varies greatly depending on age and other comorbidities, but overall this group has the highest rate of mortality rate per 100,000 as it is the most common form of PH(12). Group 2 and the other PH subgroups typically respond negatively to the therapies given in PAH. The most difficult Group 2 subtype to differentiate from PAH is PH due to HFpEF(12), due to reasons that will be outlined below. With these differences in treatment and mortality, the type of PH needs to be diagnosed accurately, early, and referred to a specialized PH center. These centers are available in most provinces, where the management of the biggest driver of mortality in PAH, i.e., development of right ventricular failure, can be monitored along with the optimal timing for lung transplantation, the ultimate treatment for advanced PAH. Misdiagnosing PH types will not only impact the patient but also the

health system, since the cost of treating PAH with 3 drug classes exceeds $200,000 per year(13, 14).

Two tests are paramount in the workup for PH: transthoracic echocardiography (TTE) and right heart catheterization (RHC). TTE usually gives the first indication that the patient has PH and is often the screening test that leads to the referral to a specialized PH center, where RHC is performed. With the TTE, features that should lead to suspicion for PH include RV dilation, reduced RV systolic function, D-shaped septum of RV pressure overload, distended inferior vena cava (IVC), shortened and notched PA acceleration time, and elevation of estimated right ventricular systolic pressure (RVSP)(6).

RHC is a vital next step in diagnosis, as it a) is the only test that can reliable separate Group 1 (where PAWP needs to be <15 mmHg) from Group 2 (where PAWP needs to be >15 mmHg(5, 6)), because TTE cannot measure PAWP, and b) because in contrast to the TTE, that can only estimate systolic PA pressure (sPAP), it can directly measure mean PA pressure (mPAP) which is required for PH diagnosis (i.e., mPAP > 20 mmHg(6)) and for the calculation of pulmonary vascular resistance (PVR=mPAP-PAWP/CO); PVR >3 Wood units is also needed to diagnose PH(5, 6). In addition, PVR is required to understand whether a patient responds to vasodilators acutely or chronically, since all current PH therapies are vasodilators and effective therapy requires a larger drop in PVR than systemic vascular resistance (SVR).

Both the TTE and RHC are subjective and operator-dependent, with a wide variability in the accuracy of measurements. TTE uses many biophysical assumptions to estimate sPAP and RV function and a slight deviation from best practice can lead to unreliable results. Similarly, RHC requires attention to detail and time, to accurately measure PAWP and CO, and the art of careful RHC varies from center to center. Thus, the possibility of test-based mistakes leading to PH misdiagnosis (or potential for missed diagnosis) is not to be ignored. Importantly, while cardiologists may be more aware of the vulnerability of TTE and RHC to mistakes and review the actual tests themselves, most physicians taking care of PH patients are non-cardiologists. Across the 17 PH clinics currently in Canada, 34 physicians are respirologists and 8 are cardiologists(15), which is a slightly larger proportion of respirologists than described in PH clinics globally (81% vs 62% described in a survey of 126 PH clinics(16)) meaning that there are less physicians experienced with TTE and RHC interpreting these results. Family physicians and internists will initiate the referral to a specialized center, along with rheumatologists, hematologists, and

respirologists. These referrals are dependent on the TTE and RHC reports for timely referral or application of guidelines directed medical therapy. We aimed to a) assess how often TTE and RHC reports encounter common mistakes, and b) how often those could lead to misdiagnosis or missed diagnosis, in the setting of a large university hospital and referral center for PH.

Methods

Study Design: In the University of Alberta PH program, all referred PH patients undergo RHC, and no PAH therapy is started without it. All patients receive PAH therapy in a goal-oriented manner, with new classes of medications typically added until the patient achieves a six-minute walk distance of >320 m, based on evidence that above this distance, the prognosis is relatively stable and more favourable(17). The program has three respirologists and one cardiologist, typical for Canada(15). Our study received appropriate approvals from the University of Alberta Ethics Board prior to initiation.

We re-reviewed the original data from all TTE's that prompted a referral to our PH center over the last 5 years, along with the initial RHC that followed the referring TTE and recorded errors in the performance and the reporting. All the TTE's and RHC's were reviewed by two cardiologists that were different from the cardiologist making the final reports. The quality of the over-read of the TTE and RHC data was assessed by comparing 20 TTEs and RHCs between the two readers, which showed a 95% agreement on the parameters that were assessed.

We then determined whether the data reported by the TTE correlated with the data reported by the RHC (the gold standard), particularly for parameters measured by both tests (e.g., right ventricular systolic pressure; RVSP). We also determined whether the data describing RV function (e.g. tricuspid annular plane systolic excursion (TAPSE) and RV size on the TTE; or cardiac output and PA pulsatility index (PAPi) in the RHC) correlate with WHO functional class assessment and the 6-minute walk tests recorded in the same timeframe with the TTE and RHC.

After recording the errors in the interpretation and reporting of the tests, we also compared the diagnosis inferred by the tests reports, to the diagnosis resulting from the correction of the TTE and RHC errors. We focused on the parameters that need to be included in a TTE report to determine the presence and severity of pulmonary hypertension as well as the parameters that need to be reported on the RHC to determine the type of pulmonary hypertension, the response to acute vasodilators (iNO), and the response to chronic PAH therapies (relative decrease in PVR over SVR), as shown in **Table 1**. For TTE, this included right atrial (RA) pressure estimation, TR velocity, TR grade, RVSP, TAPSE, RV dilation grade, and if there were other potential causes for high left atrial (LA) pressure. Each TTE was reviewed to see if poor TR jet quality was mentioned, if the TR was not checked in multiple views, if the TAPSE was recorded, if RV hypertrophy and

dilatation were quantified and if the RA pressure (RAP) was estimated appropriately using the IVC diameter and distensibility. Each RHC was reviewed for calibration quality (e.g., RAP reported as lower than -3 mmHg), whether or not respiratory variation for PAWP was recorded, or if it was recorded incorrectly (as the gold standard is to record end-expiratory values), whether or not the PAWP was the average of 3 different beats, whether the LVEDP was different from the PAWP (as without mitral stenosis, these should be the same), and whether the PAWP was greater than the dPAP (which should not be physiologically possible and clearly suggests error).

Study Population: We analyzed 263 consecutive patients with RHCs. Of these, 260 patients had a TTE within one year of the RHC (which is in keeping with usual practice, as it would be unusual for a patient to have a RHC without a TTE being performed; in these few cases without, cardiac MRIs had been performed instead). From these, 6 further patients were excluded due to severe mitral stenosis or severe pulmonic stenosis, so that this did not cloud the accuracy of comparing PAWP to left ventricular end diastolic pressure (LVEDP) or the accuracy of the TTE RVSP compared to RHC sPAP. Two patients were also excluded for severe LV systolic failure, to ensure that patients with elevated PAWPs reflected HFpEF, rather than HFrEF, giving a final sample size of 252. TTEs were performed on various echo machines and utilizing different software programs depending on the site (University hospital versus community sites), while RHCs all were interpreted using Philips IntelliVue X3® program at the University hospital.

Baseline characteristics of the population studied include 57% female, mean age of 57, mean sPAP of 56±1.6 mmHg, mean PAWP of 14±0.6 mmHg, WHO functional class 3±0.05, 6 min walk distance 343±16 m. 27% of the patients had already died at the time of analysis of the data spanning a 5-year period. Thus, this is a sick group of PH patients, typical for populations assessed in referral centers. At the time of referral, the inferred diagnosis was PAH for 111 and HFpEF for 78 patients. However, after correcting for the errors found, the inferred diagnosis often changed.

Statistical Analysis: To assess the relationship between the TTE parameters and the gold standard RHC measures, as well as to assess the reported RHC PAWP values with the gold standards of end-expiration measurement and LVEDP measurement, we reported Pearson r correlation coefficients. We also reported Pearson r correlation coefficients for CO vs WHO class and 6-minute-walk distance.

A regression analysis will show whether these values correlate, but not if they agree or the degree of agreement. For this reason, we also utilized a Bland Altman analysis to look at the agreement between the mean values of the two tests, showing the limits and potential bias of one variable compared to the gold standard variable. The Bland Altman analysis plots the difference for each paired value of the 2 tests over the mean difference of all the paired values. We also adjusted the limits of the bias based on what is considered reasonable, real-world, variation. For example, for the RVSP on TTE, it would be reasonable to expect this value to be within 10 mmHg of gold standard sPAP value (accounting for small technical differences). For the PAWP vs LVEDP, a reasonable amount of variation was within 5 mmHg.

Results

1. **TTE**

Several errors were identified in the performance of the TTEs and their interpretations and are listed in **Table 2**.

The RVSP on TTE vs sPAP on RHC was only moderately correlated (r = 0.62) with slight improvement if looking at the breakdown of those RHCs done less than two days from the TTE (r = 0.76), compared to r = 0.60 when > 2 weeks (**Figure 1**). The mean deviation of TTE RVSP from RHC sPAP (gold standard) was 18 mmHg (+/- 1) and Bland-Altman analysis showed 95% of the differences ranged from -40 to +50 mmHg. This wide degree of spread was fairly similar between the university hospital and the community sites (both with r = 0.6) as was the degree of differences on the Bland-Altman analysis (Figure 1A), suggesting that this difference was not center-specific. To explain this variation, when reviewing the TTEs to see if there was an indication of TR jet quality, we found that many studies confidently reported a value for RVSP when the TR jet quality was poor (34% of studies) (see **Table 2**). The TR jet was also not always checked in at least 3 views (12% of studies).

RV function was not always appropriately quantified. 16% of studies did not report a TAPSE value (or mention that image quality was too poor for a TAPSE to be performed). The RV function, estimated by TAPSE, was weakly correlated with sPAP (r = - 0.27), which is not surprising, as many other factors compromise RV contractility over and above increased afterload. RV thickness was usually not commented on (97% of studies with no mention of RV thickness). The degree of RV dilatation was qualitatively reported in most of the TTEs (96%), with 48% of studies reporting RV dilation.

Since the PAPi, i.e., (sPAP-dPAP)/RAP) is a RHC value often cited as a surrogate of RV function, we compared PAPi with RV function and size on TTE. There was no correlation between TAPSE and PAPi (r = 0.05, **Figure 2A**) or between the degree of RV dilation and PAPi (**Figure 2B**). PAPi also had no correlation with 6-minute walk distance (r = 0.08) or WHO-functional class (r = - 0.13) (data not shown).

RA pressure estimation was done incorrectly in 10% of TTEs (e.g. inappropriate maneuvers to assess IVC pulsatility). When comparing the RA pressure estimation on TTE to RHC, the overall correlation was poor. As this value is very dependent on loading conditions and volume status, the data was grouped into those who had TTE and RHC within two days, within 2

days to 2 weeks, and greater than two weeks of each other, and this still did not improve the correlation by a significant amount (r value of 0.5, 0.6, 0.4, respectively).

Overall, there were several instances where measurements on TTE could have led to the wrong conclusion for a patient diagnosis. Typically, a diagnosis of PH is included in the TTE report if the estimated RVSP is higher than 40 mmHg. As shown in the red box in **Figure 1A**, as well as outlined in **Table 2**, 9 patients would have been missed as having PH at all if a RHC had not also been done (i.e., the TTE reported a normal RVSP, when sPAP was truly elevated on RHC and in some cases higher than 60 or even 80 mmHg). Importantly, although these TTEs all had other features of PH (RV dilation, reduced RV function), "possible PH" or a statement that RV pressures could be underestimated was not mentioned in the report. This is important, because the falsely normal RVSP report could cause an ordering family physician or internist to not send a referral to a PH center. Several patients were also labelled as having PH on TTE, when the RHC found pressures to be normal (35 patients, **Table 2**), which has the potential for misdiagnosis but at least not a potential to prevent a referral.

2. RHC

Several errors were identified in the performance of the RHCs and their interpretations and are listed in **Table 2**.

When reviewing respiratory variation of the PAWP and the final value that was reported, 60 studies did not properly report the end-expiratory PAWP (e.g. reporting on inspiration or the mean between inspiration and expiration), while 81 studies reported a single PAWP value (as opposed to at least 3) with no specification of the respiratory cycle at all. Of those done incorrectly, 3 patients had their PAWP reported as >15 mmHg when it was truly low, which could have missed a diagnosis of PAH over HFpEF. Nine patients had their PAWP reported as <15 mmHg when it was truly higher, which could have given them an incorrect diagnosis of PAH rather than HFpEF. There could have also been more of these misclassifications, since 81 studies did not record respiratory variation at all, so it is unclear if these patients were truly in end-expiration for the value reported.

There were 120 studies that also included a left-heart catheterization, so we were able to compare the LVEDP directly to the PAWP reported (as this value should be equivalent to the LVEDP in the absence of mitral stenosis, which we had excluded from our study). The Pearson r

correlation for LVEDP vs PAWP was only 0.50 and the Bland-Altman analysis showed that 95% of the difference between LVEDP and PAWP fell between -11.3 and +14.9 mmHg (**Figure 3a**). We considered an acceptable variation from the gold standard of LVEDP would be if PAWP was ±5 mmHg, and with the reported PAWP there were 37/124 (29.8%) studies that had a difference greater than 5 mmHg. This relationship slightly improved if the LVEDP was compared to the corrected reviewed true end-expiratory PAWP (r = 0.64, 95% agreement between -9.9 to +14.3 (**Figure 3b**). When comparing the LVEDP to the PAWP in those studies that did not record any respiratory variation compared to those that did, the relationship was much worse. With no respiratory variation recorded, LVEDP vs PAWP had an r = 0.27 and Bland-Altman analysis 95% agreement between -20.8 to +12.9 mmHg, with a bias of -3.99 mmHg. With no respiratory variation recorded, 16/42 (38%) patients had >5 mmHg variation between PAWP and LVEDP. However, when respiratory variation was recorded (and interpreted properly, in retrospect) the r increased to 0.67, and in the Bland-Altman analysis 95% agreement was between -9.6 to +8.7 mmHg, with a bias of -0.43 mmHg. In that case, fewer patients (11/52, 21.1%) had >5 mmHg deviation (**Figure 3c & d**). As outlined by the red and blue boxes in Figure 3, several patients could have been misclassified if relying on the PAWP alone: the red boxes showing those patients who had a PAWP reported as >15 mmHg, but the LVEP was actually <15 mmHg (18 patients) and the blue boxes showing those who had a PAWP reported as <15 mmHg, but the LVEDP was actually >15 mmHg (11 patients). Thus, in addition to the 12 patients that were misdiagnosed based on our reanalysis of the PAWP waveforms alone, 29 more patients were misdiagnosed based on the discrepancy between PAWP and LVEDP, for a total of 41 patients (**Table 2**).

Despite these examples of inaccuracies in how the PAWP was being measured, zero PAWP saturations were performed to help confirm whether the PAWP value was correct; and there were only three studies in which they reportedly could not obtain a wedge. There were also 25 studies where the PAWP measurement did not make physiological sense, as it was a greater value than the dPAP, definitely pointing to an error.

CO measurements were done in all RHC studies, however the majority (86%; 217 studies) were done using the indirect Fick's method alone. Nine studies used thermodilution alone, and 26 studies used both indirect Fick's method and thermodilution. In the small subset of studies that used both, we were able to compare CO measurements, and they did not correlate well, with an r

value of 0.44 for Fick vs thermodilution and Bland-Altman analysis showed that 95% agreement range from -2.2 to +2.6 (**Figure 4**).

There were several instances where the PVR value changed from >3 Wood units (in keeping with PH) to <3 Wood units, depending on which measure of CO was used. When correcting for PAWP (i.e. by using the correct end-expiratory PAWP to calculate PVR using a Fick CO) the PVR went from <3 WU to >3 Wood units 33 times and from >3 WU to <3 WU 8 times. When using the correct end-expiratory PAWP and a thermodilution CO, 3 out of 35 studies changed PVR category. Therefore, the accuracy of the PAWP and the method of CO measurement can both have a large impact on the PVR calculations and thus have clinical implications.

Due to restrictions from the COVID-19 pandemic that took place during our study period, there were 185 clinical assessments closely synchronized with TTE and RHC assessments that recorded WHO functional status and 61 that recorded 6-minute walking distance. The 6-minute walking distance and WHO functional class had an expected negative correlation to each other (i.e.: that as the WHO functional class increased from I to IV, the 6-minute walking distance worsened), with an r value of - 0.74. WHO functional class and 6-minute walking distance both did not correlate with Fick CO values but there was a moderate correlation between with thermodilution CO values, with an r value of -0.46 and 0.45 respectively (**Figure 5**). There was also no correlation between CO and TAPSE or grade of RV dilation (not shown).

Discussion

Here we report that preventable errors can commonly occur during the performance and reporting of TTEs and RHCs, even in specialized referral centers, with a significant impact in the care of PH patients as well as in the health care system. Overall, these errors affected the correct diagnosis in 34 % of the referred PH patients (**Table 2**). The implications of these errors include:

1) Delays in the diagnosis of possible PH at the primary care level, which in turn may cause delays in the referral of patients to specialized PH centers. The delay of PAH diagnosis is known to adversely affect the survival of patients and the optimal management of RV failure(4).

2) Misdiagnosis of PAH versus PH due to HFpEF. Those two conditions have significant differences in prognosis and treatment. At the specialist level, this is particularly relevant after the

recent introduction of therapies like SGLT inhibitors that can be of significant benefit in HFpEF and prevent RV failure(18). As **Table 2** shows >16% of patients can suffer from misdiagnosis.

3) At the level of non-cardiology specialists that manage PH patients (like respirologists, rheumatologists, and hematologists), the mistakes in PVR can lead to misinterpretation of the response of PH patients to PAH-therapies, since they are all vasodilators aiming for a preferential decrease in PVR over SVR. This may lead to wrong timing and application of guideline-directed medical therapies.

4) Considering the huge costs of PAH therapies, these errors can have a significant impact on the health care system, in addition to the application of the wrong medications (and thus their side effects) in the wrong patient population.

The majority of the TTE errors can be fixed by adhering to the standard practice of echocardiography and by the understanding of physicians preparing the reports, that a careful statement of both the quality of the RVSP estimation as well as the description of the RV size and function, could have important implications on whether a primary care physician will follow up with a referral to a specialized PH referral center. Similarly, the majority of the RHC errors can be fixed by the understanding of the performing cariologists that confirming an appropriate PAWP position (with an O_2 saturation check), recording PAWP at the end-expiratory position and for at least 3 heart beats is critical, since deviation of as little as 5mmmHg can have significant prognostic and therapeutic implications for patients diagnosed with PAH versus PH secondary to HFpEF. Another conclusion from our work is that the indirect Fick method of calculating CO does not appear to correlate with the functional status of the patients or the TTE-measured function of the RV. Reasons for this have been well described(19), and rest with the several levels of estimation in the equation ($CO = VO_2/Ca\text{-}Cv$). VO_2 is estimated by using BSA multiplied by a constant. In our centre, a constant value of 133 is used (which is different from most studies, which use the Dehmer formula of 125 x BSA(20)). The 133 value is a proprietary default setting by the company supplying our catheterization program (Philips), and as far as we can see, has not been studied or validated in published literature. Studies on the accuracy of these estimation methods show that other formulas may have somewhat better accuracy (like the LaFarge & Miettinen formula(21)) so one area for potential improvement a change to the VO_2 estimation formula. Despite these

known limitations, the vast majority of studies used the Fick method, presumably in an attempt to save time.

A limitation of this study is that the comparison of TTE with RHC should ideally be attempted if the two studies are performed simultaneously, perhaps as part of a research protocol. However, this is not realistic in clinical practice. In fact, a strength of our study is that it reflects routine clinical practice, increasing its relevance to both non-specialists and specialists dealing with PH. Another limitation is that the use of vasodilatory medication prior to initiation of left heart catheterization measurements (like LVEDP) was not easily obtainable, and therefore differences from PAWP to LVEDP may be reflective from an overall drop in pressure – however we do anticipate that this change should be still small and varying >5 mmHg (and in particular >10 mmHg) would not be expected.

Our results should alert echocardiography and catheterization laboratories staff as well. Although we focused on studies of patients referred to our center with PH as a running diagnosis, the same errors we found apply to all the patients undergoing these tests, potentially significantly increasing the impact of these errors, extending it non-PH populations.

Clinical Perspectives

For the primary care physician ordering the initial TTE, it is important to realize the vulnerabilities of TTE. If a patient has symptoms of right-heart failure, like peripheral edema, ascites, or hepatosplenomegaly in addition to decreased functional capacity, the threshold for referral to a PH center should be low. For example, if indirect signs of PH appear in the report (like RV dilatation or RV hypertrophy), the patient should be referred even if the diagnosis of PH is not included in the conclusions or the estimated RVSP appears to be only mildly elevated.

For the PH expert, it is important to ask the right questions when referring a patient for RHC and to emphasize why the accuracy of those values will be important. If the clinical suspicion is high and the quality of TTE and RHC suboptimal, the testing may need to be repeated. For the cardiologists it is important to be reminded that although often RHC is considered a simple skill (compared to a high-risk intervention for example) mistakes in tasks as simple as recording a PAWP are common and can make a huge difference for a patient. Although not all our RHC performing cardiologists made such errors, some of the ones that consistently made them were senior and experienced interventional cardiologists.

Finally, for the health care system at large, this should trigger conduction of quality improvement projects in echocardiography and catheterization laboratories. Thankfully, there are studies showing that these errors can be remedied through quality improvement programs. One study at an academic hospital in Germany showed that with coaching, RHC sPAP to TTE RVSP obtained an r value of 0.87(22). Another study of two large academic hospitals in Delaware, USA showed that after a teaching intervention, their correlation between several TTE and RHC parameters improved significantly. For example, TTE RVSP vs RHC RVSP had r values improve from 0.30 to 0.77 and TTE reports of diastolic dysfunction vs LVEDP had r values improve from 0.09 to 0.62)(23). In another study, simply changing from measuring "the chin" instead of "the beard" for over-gained continuous wave tricuspid regurgitation Doppler signals, significantly improved correlation of TTE with RHC(24). One study comparing PAWP measures with and without doing PAWP saturations found that only 50% of initial PAWP measurements were truly occlusive, and that when wedging was repeated until a true PAWP saturation was obtained, the average PAWP was often at least 5 mmHg lower, which resulted in reclassification of PH group in 10% of those patients(25).

With repetition and reinforcement, any suggested changes can become the standard of care and these improvements could be applicable to centers across the country which may have encounter similar errors, as this problem certainly exists outside our institution. The goal of the project is to bring attention to these areas for improvement, so that operators and readers in the catheterization and echo labs can have more consistency across the country and to hopefully improve patient care and quality of life through accurate diagnosis and management.

References

1. Gall H, Felix JF, Schneck FK, Milger K, Sommer N, Voswinckel R, et al. The Giessen Pulmonary Hypertension Registry: Survival in pulmonary hypertension subgroups. The Journal of Heart and Lung Transplantation. 2017;36(9):957-67.

2. Benza RL, Miller DP, Barst RJ, Badesch DB, Frost AE, Mcgoon MD. An Evaluation of Long-term Survival From Time of Diagnosis in Pulmonary Arterial Hypertension From the REVEAL Registry. Chest. 2012;142(2):448-56.

3. Chang KY, Duval S, Badesch DB, Bull TM, Chakinala MM, De Marco T, et al. Mortality in Pulmonary Arterial Hypertension in the Modern Era: Early Insights From the Pulmonary Hypertension Association Registry. Journal of the American Heart Association. 2022;11(9).

4. Khou V, Anderson JJ, Strange G, Corrigan C, Collins N, Celermajer DS, et al. Diagnostic delay in pulmonary arterial hypertension: Insights from the Australian and New Zealand pulmonary hypertension registry. Respirology. 2020;25(8):863-71.

5. Simonneau G, Montani D, Celermajer DS, Denton CP, Gatzoulis MA, Krowka M, et al. Haemodynamic definitions and updated clinical classification of pulmonary hypertension. European Respiratory Journal. 2019;53(1):1801913.

6. Humbert M, Kovacs G, Hoeper MM, Badagliacca R, Berger RMF, Brida M, et al. 2022 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: Developed by the task force for the diagnosis and treatment of pulmonary hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS). Endorsed by the International Society for Heart and Lung Transplantation (ISHLT) and the European Reference Network on rare respiratory diseases (ERN-LUNG). European Heart Journal. 2022;43(38):3618-731.

7. Paulin R, Michelakis ED. The Metabolic Theory of Pulmonary Arterial Hypertension. Circulation Research. 2014;115(1):148-64.

8. Humbert M, Mclaughlin V, Gibbs JSR, Gomberg-Maitland M, Hoeper MM, Preston IR, et al. Sotatercept for the Treatment of Pulmonary Arterial Hypertension. New England Journal of Medicine. 2021;384(13):1204-15.

9. Gillies H, Niven R, Dake BT, Chakinala MM, Feldman JP, Hill NS, et al. AV-101, a novel inhaled dry-powder formulation of imatinib, in healthy adult participants: a phase 1 single and multiple ascending dose study. ERJ Open Research. 2023;9(2):00433-2022.

10. Ezekowitz JA, O'Meara E, Mcdonald MA, Abrams H, Chan M, Ducharme A, et al. 2017 Comprehensive Update of the Canadian Cardiovascular Society Guidelines for the Management of Heart Failure. CJC. 2017;33(11):1342-433.

11. Mcgetterick M, Peacock A. Group 3 pulmonary hypertension: Challenges and opportunities. Global Cardiology Science and Practice. 2020;2020(1).

12. Kang M, Hart CM, Kempker JA, Veeraraghavan S, Trammell AW. Pulmonary hypertension mortality trends in United States 1999–2019. Annals of Epidemiology. 2022;75:47-52.

13. Sikirica M, Iorga SR, Bancroft T, Potash J. The economic burden of pulmonary arterial hypertension (PAH) in the US on payers and patients. BMC Health Services Research. 2014;14(1).

14. CADTH. Macitentan (Opsumit): For Long-Term Treatment of Pulmonary Arterial Hypertension [Internet]. Ottawa (ON): Canadian Agency for Drugs and Technologies in Health; 2015 Jul. Table 1, Cost Comparison Table for Drugs Used for the Treatment of Pulmonary Arterial Hypertension 2015 [Available from: Available from: https://www.ncbi.nlm.nih.gov/books/NBK349251/table/T30/.

15. PHA. Find a PH Centre https://phacanada.ca/Living-with-PH/PHCentres: PHA Canada; [

16. Doyle-Cox C, Nicholson G, Stewart T, Gin-Sing W. Current organization of specialist pulmonary hypertension clinics: results of an international survey. Pulmonary Circulation. 2019;9(2):1-10.

17. Maron BA. Revised Definition of Pulmonary Hypertension and Approach to Management: A Clinical Primer. Journal of the American Heart Association. 2023;12(8).

18. Anker SD, Butler J, Filippatos G, Ferreira JP, Bocchi E, Böhm M, et al. Empagliflozin in Heart Failure with a Preserved Ejection Fraction. New England Journal of Medicine. 2021;385(16):1451-61.

19. Fares WH, Blanchard SK, Stouffer GA, Chang PP, Rosamond WD, Ford HJ, et al. Thermodilution and Fick Cardiac Outputs Differ: Impact on Pulmonary Hypertension Evaluation. Canadian Respiratory Journal. 2012;19(4):261-6.

20. Khirfan G, Ahmed MK, Almaaitah S, Almoushref A, Agmy GM, Dweik RA, et al. Comparison of Different Methods to Estimate Cardiac Index in Pulmonary Arterial Hypertension. Circulation. 2019;140(8):705-7.

21. Chase PJ, Davis PG, Wideman L, Starnes JW, Schulz MR, Bensimhon DR. Comparison of Estimations Versus Measured Oxygen Consumption at Rest in Patients With Heart Failure and Reduced Ejection Fraction Who Underwent Right-Sided Heart Catheterization. The American Journal of Cardiology. 2015;116(11):1724-30.

22. Greiner S, Jud A, Aurich M, Hess A, Hilbel T, Hardt S, et al. Reliability of Noninvasive Assessment of Systolic Pulmonary Artery Pressure by Doppler Echocardiography Compared to Right Heart Catheterization: Analysis in a Large Patient Population. Journal of the American Heart Association. 2014;3(4):e001103-e.

23. Fanari Z, Choudhry UI, Reddy VK, Eze-Nliam C, Hammami S, Kolm P, et al. The Value of Quality Improvement Process in the Detection and Correction of Common Errors in Echocardiographic Hemodynamic Parameters in a Busy Echocardiography Laboratory. Echocardiography. 2015;32(12):1778-89.

24. Kyranis SJ, Latona J, Platts D, Kelly N, Savage M, Brown M, et al. Improving the echocardiographic assessment of pulmonary pressure using the tricuspid regurgitant signal-The “chin” vs the “beard”. Echocardiography. 2018;35(8):1085-96.

25. Viray MC, Bonno EL, Gabrielle ND, Maron BA, Atkins J, Amoroso NS, et al. Role of Pulmonary Artery Wedge Pressure Saturation During Right Heart Catheterization. Circulation: Heart Failure. 2020;13(11).