



Article

# Diagnostic Errors Induced by a Wrong a Priori Diagnosis: A Prospective Randomized Simulator-Based Trial

Felix M.L. Meyer <sup>1</sup>, Mark G. Filipovic <sup>2</sup> , Gianmarco M. Balestra <sup>3</sup>, Kai Tisljar <sup>3</sup>, Timur Sellmann <sup>4,5</sup> and Stephan Marsch <sup>3,\*</sup>

<sup>1</sup> Department of Intensive Care, Kantonsspital Luzern, 6000 Luzern, Switzerland; felix.meyer@luks.ch  
<sup>2</sup> Institute of Anesthesiology, Kantonsspital Winterthur, 8400 Winterthur, Switzerland; mark.filipovic@ksw.ch  
<sup>3</sup> Department of Intensive Care, University of Basel Hospital, 4031 Basel, Switzerland; gianmarco.balestra@usb.ch (G.M.B.); kai.tisljar@usb.ch (K.T.)  
<sup>4</sup> Department of Anaesthesiology, Witten/Herdecke University, 58455 Witten, Germany; t.sellmann@bethesda.de  
<sup>5</sup> Department of Anaesthesiology, Bethesda Hospital, 47053 Duisburg, Germany  
\* Correspondence: stephan.marsch@usb.ch



**Citation:** Meyer, F.M.L.; Filipovic, M.G.; Balestra, G.M.; Tisljar, K.; Sellmann, T.; Marsch, S. Diagnostic Errors Induced by a Wrong a Priori Diagnosis: A Prospective Randomized Simulator-Based Trial. *J. Clin. Med.* **2021**, *10*, 826. <https://doi.org/10.3390/jcm10040826>

Academic Editor: John G. Kellett

Received: 22 January 2021

Accepted: 14 February 2021

Published: 18 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Preventive strategies against diagnostic errors require the knowledge of underlying mechanisms. We examined the effects of a wrong a priori diagnosis on diagnostic accuracy of a focussed assessment in an acute myocardial infarction scenario. One-hundred-and-fifty-six medical students (cohort 1) were randomized to three study arms differing in the a priori diagnosis revealed: no diagnosis (control group), myocardial infarction (correct diagnosis group), and pulmonary embolism (wrong diagnosis group). Forty-four physicians (cohort 2) were randomized to the control group and the wrong diagnosis group. Primary endpoint was the participants' final presumptive diagnosis. Among students, the correct diagnosis of an acute myocardial infarction was made by 48/52 (92%) in the control group, 49/52 (94%) in the correct diagnosis group, and 14/52 (27%) in the wrong diagnosis group ( $p < 0.001$  vs. both other groups). Among physicians, the correct diagnosis was made by 20/21 (95%) in the control group and 15/23 (65%) in the wrong diagnosis group ( $p = 0.023$ ). In the wrong diagnosis group, 31/52 (60%) students and 6/23 (19%) physicians indicated their initially given wrong a priori diagnosis pulmonary embolism as final diagnosis. A wrong a priori diagnosis significantly increases the likelihood of a diagnostic error during a subsequent patient encounter.

**Keywords:** diagnostic error; randomized controlled trial; myocardial infarction; pulmonary embolism; simulation

## 1. Introduction/Background

Compared to a vast variety of medical errors, diagnostic errors have only recently gained political and academic attention [1–4]. The 2015 report “Improving diagnosis in health care” of the Institute of Medicine defines a diagnostic error as failure to establish an accurate and timely explanation of the patient’s health problem or communicating that explanation to the patient [1]. Because of their central role in medical decision making and strong influence on subsequent procedures and therapy, faults in diagnosis can trigger far-reaching consequences [5,6]. Among malpractice claims, diagnostic errors appear to be the most common, most costly, and most dangerous of medical mistakes [7]. The true incidence of diagnostic errors is difficult to establish. Based on the currently available data, a fair estimate of the prevalence of diagnostic errors is approximatively 5% and most people will likely be affected by at least one diagnostic error during their lifetime [1,8–10].

The current knowledge on diagnostic errors mainly originates from retrospective error analysis and observational studies, both prone to hindsight bias [11]. In addition, not all errors can be detected retrospectively. Indeed, available data from prospective

studies suggest that the incidence of diagnostic errors and their consequences like inappropriate or even dangerous management may be significantly higher than estimated from observational studies [6,12–15].

Patients with acute and severe illness presenting at emergency departments are generally triaged during a first patient–provider encounter. The rapid evolution of their disease and the limited opportunities to detect and correct errors in the settings of acute care makes these patients particularly vulnerable to harm related to a delayed or wrong diagnosis [16]. Thus, the diagnostic accuracy in the first patient–provider encounter is paramount for patients' safety. Indeed, the implementation of systematic cross-checking between emergency physicians was associated with a significant reduction in adverse events [17].

Patients presenting to emergency departments often hold a suspected diagnosis of a pre-treating physician. Suspected diagnoses may be reported by a variety of sources like other doctors, paramedics, triage nurses, or from the patients themselves, and may be correct or wrong. To the best of our knowledge, there are no data on the influence of an a priori diagnosis on potential diagnostic errors during a subsequent patient encounter.

Accordingly, the aim of our study was to test the hypothesis that a wrong a priori diagnosis of a referring physician will increase the likelihood of wrong final diagnoses.

## 2. Methods

### 2.1. Study Design

This is an investigator-initiated prospective randomized controlled single-blind trial, the participants not being aware of the purpose of the study. The trial (ClinicalTrials.gov NCT04659265) was approved by the regional ethical committee (EKNZ 85/04), and all participants gave written informed consent.

### 2.2. Participants

Two cohorts were studied: In cohort 1, medical students in their 4th year at the University of Basel were offered voluntary workshops at our institution. The present study was conducted as part of these workshops. Second, the results of the first cohort prompted us to perform a study with physicians (residents and fellows in internal medicine and intensive care medicine) participating in simulator workshops at our institution (cohort 2).

At the beginning of the workshop, participants were asked to fill in a questionnaire. The questionnaire for the medical students included (a) the Big Five personality traits (neuroticism, extraversion, openness, agreeableness, conscientiousness) [18]; (b) Rosenberg's self-esteem scale [19]; (c) the task to diagnose the cardiac rhythm in 12 one-lead ECG rhythms strips (each correct diagnosis was counted as one point resulting in a maximum of 12 attainable points); (d) the task to list the typical symptoms of an acute myocardial infarction in bullet point form (each of the following items were counted as one point resulting in a maximum of six attainable points: chest pain; radiation of the pain; pain described as constricting, compressing, or squeezing; at least one accompanied vegetative symptom; dyspnoea; and fear); and (e) the task to list the typical signs symptoms of four further medical emergencies (anaphylactic shock, trauma management, cardiac arrest, and syncope) in bullet point form (answers were coded using a predefined checklist). The questionnaire for the physicians included the Big Five personality traits and Rosenberg's self-esteem scale.

### 2.3. Simulator and Scenario

The study was conducted using a patient simulator. All participants received a standardised instruction of the patient simulator (human patient simulator, SimMan<sup>®</sup>, Laerdal Medical AS, Stavanger, Norway). This simulator's features include among others, palpable pulses, spontaneous breathing with visible thoracic excursions, spontaneous movements of the eye lids, and a speaker in the mannequin's head that broadcasts the voice of a remote operator, allowing for verbal interaction with the participants. For the present trial, the operator was an experienced physician (G.B., K.T., or S.M.)

For the study scenario, participants were given the role of a resident-on-call in the emergency department and were instructed by a senior physician (confederate) to triage a newly arrived patient. Upon entering the simulator room, the participants encountered the patient (simulator) lying on a bed and dressed in a hospital gown. Participants were assisted by a confederate nurse, which was instructed to display a helpful manner, to act only on command, but not to interfere in any way with the assessment and history taking.

#### 2.4. Randomization and Intervention

All participants performed the scenario alone and faced a patient with a simulated acute myocardial infarction. In cohort 1, the medical students were randomized (computer-generated numbers) 1:1:1 to three versions of the scenario. The three versions differed only in the reply of the patient to the initial question about the reason for his visit. The patient stated that he had phoned his family physician because of acute chest pain and dyspnoea. The family physician had recommended urgent visit of the emergency department because he judged the situation as (1) medical emergency (control group with no a priori diagnosis); (2) acute myocardial infarction (correct a priori diagnosis group); or, (3) acute pulmonary embolism (wrong a priori diagnosis group). Apart from this randomly allocated initial information, the scenario was identical in all groups: until the end of the scenario, the patient only spoke when asked a question and all patient's answers were given according to a pre-scripted answer check-list (Table 1).

**Table 1.** Checklist of the patient's answers to questions asked by the participants during history taking.

Questions of the Participants	Answers of the Patient
<b>Exploration of chest pain (chest pain bloc)</b>	
Localisation	In the middle of the chest behind the breastbone
Quality	Pressing and constricting
Radiation	Yes, to the left elbow and neck
Severity	Severe. When a pain score is asked: 8 out of 10
Aggravating or relieving factors	None
Previous experience	Never experienced something similar before
Duration	Continuous pain over the last 30 min
Circumstances	Sudden onset while sitting at the desk working. When asked: no preceding physical activity or unusual stress
<b>Associated symptoms</b>	
Nausea or vomiting	Slight Nausea, no vomiting
Sweating	Initially heavy sweating, that now has ceased
Fear	Initially strong fear of death, now frightened
Dyspnoea	Slight shortness of breath
<b>Cardiovascular risk factors</b>	
Smoking	One package cigarettes a day for the last 30 years
Hypertension	No. Blood pressure normal, according to GP
Diabetes	No
Blood fats	Normal, according to GP
Family history	Mother died due to a stroke with 70 years; father 82 years old with dementia; two healthy brothers

Table 1. Cont.

Questions of the Participants	Answers of the Patient
<b>Previous illness and medication</b>	
Prior diseases	None
Prior hospitalisations	Appendectomy when 16 years old
Current medication	None
Allergies	Not known
Last visit with GP	Normal “check-up” one year ago
<b>Social history</b>	
Age	49 years old
Next of kin	Married, no children
Profession/occupation	Accountant in banking
Recreation activities/hobbies	Cycling, fitness centre, traveling
<b>Exploration for pulmonary embolism</b>	
Cough	None
Haemoptoe	No
Chest pain aggravated by breathing	No relation of chest pain with breathing activity
Painful or swollen leg(s)	No
Previous thrombotic event	No
Immobilisation	No. When specifically asked: no trauma, no operation, no recent travel
Risk factors for thrombotic event	When specifically asked: patient not aware of cancer, coagulopathy, coagulopathy in family

In cohort 2, physicians were randomized (computer-generated numbers) 1:1 to the control group (no a priori diagnosis) and the wrong a priori diagnosis group.

All participants of the groups with no a priori diagnosis and correct a priori diagnosis were given 4 min assessment time. Participants of the group with the wrong a priori diagnosis were given one additional minute (i.e., 5 min assessment time) to compensate for probing the given diagnosis. At the end of the study time, the senior physician entered the room and announced that he required the participant to temporarily leave his/her patient to help with another more urgent case. While still in the room, the participants were then asked on their presumptive diagnosis and which measures, if any, the nurse should take during their temporary absence during the next couple of minutes.

### 2.5. Data Analysis

Data analysis was performed using video recordings obtained during simulations. Video recordings were analysed independently by two trained raters, who noted all questions asked and all measures taken by the participant. Inter-observer differences were solved by jointly re-watching the video recordings.

### 2.6. Statistics

The primary endpoint was a correct final presumptive diagnosis of an acute myocardial infarction or acute coronary syndrome when asked at the end of the scenario by the senior physician or when previously revealed to the patient and/or the nurse. Mentioning the correct diagnosis as possible differential diagnosis was rated as correct diagnosis. A co-primary endpoint was the composite endpoint of a correct final presumptive diagnosis OR ordering a 12-lead ECG, the rationale being that a timely 12-lead ECG will most

likely prevent harm to a patient not correctly triaged as suffering from acute myocardial infarction.

Secondary endpoints were the number and kind of questions asked during history taking; probing of the a priori diagnosis, defined as asking for symptoms or performing examinations specific for the given a priori diagnosis; measures taken during assessment; and the results of the questionnaire. Results of the knowledge part of the questionnaire are reported as percentages of the attainable maximum.

A 20% difference in the primary outcome was considered to be of medical relevance. A power analysis revealed that for the cohort of medical students, approximately 150 participants (50 per group) and for the cohort of physicians approximately 80 (40 per group) had to be included to detect this difference with a power of 0.9 at the 0.05 significance level. A single interim analysis was scheduled after recruitment of half the planned cohort of physicians with two-sided hypothesis testing with boundaries allowing for early stopping either in favour of the null or of the alternative [20,21].

Statistical analysis was performed by SPSS (version 25; IBM, Armonk, NY, USA). Non-parametric ANOVA, Mann–Whitney test, Fisher’s exact test, Univariate and Multivariate regression analysis of variables associated with diagnostic accuracy, and Cohen’s kappa were used as appropriate. A  $p < 0.05$  was considered to represent statistical significance. Data are reported as median [Interquartile range {IQR}] unless otherwise stated.

### 3. Results

#### 3.1. Medical Students

Out of 316 students offered the possibility, 156 (49%) participated in the voluntary workshops, were included in the trial, and completed the trial as intended. Fifty-two participants (31 females) were randomised to the control group, 52 (34 females) to the correct a priori diagnosis group, and 52 (30 females) to the wrong a priori diagnosis group. No protocol violations occurred and video-recordings of all participants were of sufficient quality for analysis. Complete questionnaire data were available from all participants. Thus, data of all 156 participants were analysed and reported. Groups were well balanced for knowledge and personality (Table 2). Cohen’s kappa for inter-rater reliability was 0.96.

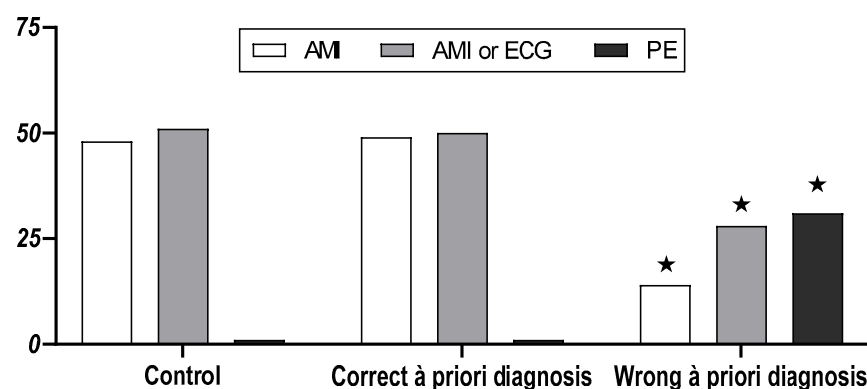
**Table 2.** Questionnaire data of the student participants.

	Control Group ( <i>n</i> = 52)	Correct a Priori Diagnosis Group ( <i>n</i> = 52)	Wrong a Priori Diagnosis Group ( <i>n</i> = 52)
Female:male	31:21	34:18	30:22
AMI knowledge %	50 [33–67]	50 [33–67]	50 [50–67]
Emergencies knowledge %	48 [37–60]	42 [27–58]	44 [37–54]
EKG knowledge %	42 [25–54]	33 [29–50]	42 [25–58]
Self-esteem (Rosenberg)	23 [20–26]	24 [20–26]	22 [17–24]
Neuroticism	2.8 [2.2–3.3]	2.9 [2.2–3.3]	2.9 [2.5–3.5]
Extraversion	4.3 [3.7–4.8]	4.2 [3.8–4.8]	4.2 [3.7–4.7]
Openness	4.0 [3.7–4.6]	4.5 [3.9–4.8]	4.2 [3.8–4.6]
Agreeableness	4.7 [4.3–5.0]	4.7 [4.5–5.0]	4.7 [4.3–5.0]
Conscientiousness	4.7 [3.8–4.8]	4.5 [3.9–4.8]	4.3 [3.8–5.0]

Data are medians [Interquartile range {IQR}]; AMI = acute myocardial infarction. ECG = Electrocardiogram. Data on knowledge are expressed as percentages of the maximum attainable result. Personality traits (neuroticism, extraversion, openness, agreeableness, and conscientiousness) are based on 6 points Likert scales. There are no significant differences between the groups.

Primary endpoint: The correct final diagnosis was made by 48/52 (92%) participants in the control group, by 49/52 (94%) participants in the group with the correct a priori

diagnosis, and by 14/52 (27%) participants in the group with the wrong a priori diagnosis ( $p < 0.001$  vs. both other groups). In the group with the wrong a priori diagnosis, 31/52 (60%;  $p < 0.001$  vs. both other groups) indicated pulmonary embolism, i.e., the initially given wrong diagnosis, as their final diagnosis compared to 1/51 (2%) in the control group and 1/51 (2%) in the correct a priori diagnosis group (Figure 1). The composite primary endpoint of a correct final diagnosis or ordering a 12-lead ECG was reached by 51/52 (98%) in the control group, by 50/52 (96%) in the group with the correct a priori diagnosis, and by 28/52 (54%) in the group with wrong a priori diagnosis ( $p < 0.001$  vs. both other groups).



**Figure 1.** Final presumptive diagnosis of medical student participants after a focussed assessment of a patient with AMI in the control group ( $n = 52$ ), the correct a priori diagnosis group ( $n = 52$ ), and the wrong a priori diagnosis group ( $n = 52$ ). AMI = diagnosis acute myocardial infarction; AMI or ECG = diagnosis acute myocardial infarction OR ordering of a 12-lead ECG; PE = pulmonary embolism. ★ =  $p < 0.05$  vs. both other groups.

Secondary endpoints: Despite having 20% more time available, we observed a similar number of questions asked during history taking in the wrong a priori diagnosis group and the control group (7 [IQR 5–10] vs. 7 [IQR 5–10];  $p = 0.97$ ). The number of questions asked in the correct a priori diagnosis group 5 [IQR 3–9] was significantly lower than in both the control group ( $p = 0.02$ ) and in the wrong diagnosis group ( $p = 0.032$ ). The number of questions asked per minute of assessment time was significantly higher in the control group (1.8 [IQR 1.3–2.4]) than in both the groups with correct (1.3 [IQR 0.8–2.1];  $p = 0.02$ ) and wrong (1.4 [IQR 1.0–1.9];  $p = 0.03$ ) a priori diagnoses. Diagnostic or therapeutic measures taken during the triage scenario are displayed in Table 3. Note that measures taken concurred with the presumptive diagnosis.

Comparisons within the wrong a priori diagnosis group: Participants with a correct and wrong final diagnosis did not differ in gender, knowledge, self-esteem, or personality traits (Table 4).

Moreover, there was no difference in the overall number of questions asked during history taking between participants of the wrong a priori diagnosis group making the correct or false final diagnosis (7 [IQR 4–10] vs. 7 [IQR 6–10];  $p = 0.42$ ). There was, however, a significant difference in the number of questions asked in the thematic bloc relating to the exploration of chest pain (2 [IQR 1–5] vs. 4 [IQR 3–5];  $p = 0.016$ ), but not in questions asked in any of the other thematic blocs. Univariate analysis revealed a significant correlation of making the correct presumptive diagnosis with asking of the quality of chest pain (beta 0.47;  $p < 0.0001$ ), the radiation of chest pain (beta 0.47;  $p < 0.0001$ ), and having had similar symptoms before (beta 0.28;  $p < 0.045$ ). In addition, there was a non-significant trend for a correlation with asking of the localization of chest pain (beta 0.26;  $p < 0.066$ ) while none of the other questions reached a  $p$  value of  $< 0.10$ . In multivariate analysis, asking about both the quality of chest pain (beta 0.30;  $p = 0.049$ ) and radiation of chest pain (beta 0.30;  $p = 0.049$ ) emerged as independent predictors of making a correct presumptive diagnosis. Interestingly, the knowledge that the quality of chest pain and radiation are typical symptoms of an AMI, as evidenced by the questionnaire, was not significantly

correlated with the likelihood of asking the corresponding questions during history taking (correlation coefficients  $-0.01$  and  $0.09$ ,  $p$  values  $0.96$  and  $0.54$  respectively).

**Table 3.** Measures taken by the student participants.

	Control Group ( <i>n</i> = 52)	Correct a Priori Diagnosis Group ( <i>n</i> = 52)	Wrong a Priori Diagnosis Group ( <i>n</i> = 52)
Number of measures taken ( <i>n</i> )	4.5 [3.0–6.0]	5.0 [3.0–5.5]	5.0 [4.0–5.0]
Supplemental oxygen ( <i>n</i> ; %)	28/52 (54%)	29/52 (56%)	16/52 (31%) *¶
Cue ECG monitoring ( <i>n</i> ; %)	25/52 (48%)	30/52 (58%)	28/52 (54%)
Blood pressure and heart rate measured ( <i>n</i> , %)	43/52 (83%)	39/52 (75%)	35/52 (67%)
Oxygen saturation measured ( <i>n</i> ; %)	3/52 (6%)	1/52 (2%)	3/52 (6%)
Chest auscultation ( <i>n</i> ; %)	32/52 (62%)	25/52 (48%)	42/52 (81%) *¶
Examination of legs ( <i>n</i> ; %)	0/52 (0%)	0/52 (0%)	9/52 (17%) *¶
Morphium given ( <i>n</i> ; %)	32/52 (62%)	29/52 (56%)	12/52 (23%) *¶
Nitro-glycerin given ( <i>n</i> ; %)	25/52 (48%)	22/52 (42%)	4/52 (8%) *¶
Aspirin given ( <i>n</i> ; %)	9/52 (17%)	13/52 (25%)	1/52 (2%) *¶
Heparin given ( <i>n</i> ; %)	0/52 (0%)	4/52 (8%)	3/52 (6%)
Chest Radiograph ordered ( <i>n</i> ; %)	0/52 (0%)	0/52 (0%)	9/52 (17%) *¶
Chest CT ordered ( <i>n</i> ; %)	0/52 (0%)	0/52 (0%)	9/52 (17%) *¶
Blood sample taken ( <i>n</i> ; %)	24/52 (46%)	25/52 (48%)	43/52 (83%) *¶
Troponin requested ( <i>n</i> ; %)	14/52 (27%)	16/52 (31%)	4/52 (8%) *¶
D-Dimer requested ( <i>n</i> ; %)	0/52 (0%)	3/52 (6%)	22/52 (42%) *¶

Data in the first row are medians [Interquartile range {IQR}]; Digits in the following rows indicate the number of participants of the given group taking the specified measure. \* =  $p < 0.05$  vs. control group; ¶ =  $p < 0.05$  vs. correct diagnosis group.

**Probing:** In the wrong diagnosis group, we observed only limited probing of the a priori diagnosis, defined as asking for specific symptoms or risk factors of pulmonary embolism, with no significant difference between participants making the correct or wrong final diagnosis (1 [IQR 1–2] vs. 1 [IQR 0–1];  $p = 0.20$ ).

Likewise, compared to the control group, we observed in the correct a priori diagnosis group significantly fewer probing questions, i.e., questions relating to chest pain symptoms (3 [IQR 2–4] vs. 4 [IQR 3–5];  $p = 0.040$ ) and risk factors of coronary disease (0 [IQR 0–2] vs. 1 [IQR 0–2];  $p = 0.037$ ), while there was no difference between these groups in the number of questions in any of the other thematic blocs.

**Table 4.** Comparisons of student participants randomized to the wrong diagnosis group.

	Final Diagnosis Acute Coronary Syndrome ( <i>n</i> = 21)	Final Diagnosis Pulmonary Embolism ( <i>n</i> = 31)
Female:male ( <i>n</i> )	13:8	17:14
AMI knowledge %	67 [50–67]	50 [50–67]
Emergencies knowledge %	45 [30–53]	42 [39–57]
ECG knowledge %	33 [25–58]	42 [33–58]
Self-esteem (Rosenberg)	22 [20–23]	22 [16–24]
Neuroticism	2.8 [2.7–3.2]	3.0 [2.5–3.8]
Extraversion	4.3 [3.8–4.7]	4.2 [3.7–4.5]
Openness	4.3 [3.8–4.7]	4.0 [3.7–4.5]
Agreeableness	4.7 [4.3–5.0]	4.7 [4.3–5.0]
Conscientiousness	4.5 [4.0–5.0]	4.2 [3.7–5.0]
Number of measures taken ( <i>n</i> ; %)	5.0 [4.0–5.0]	5.0 [4.0–6.0]

Data are medians [Interquartile range (IQR)]; AMI = acute myocardial infarction. Data on knowledge are expressed as percentages of the maximum attainable result. Personality traits (neuroticism, extraversion, openness, agreeableness, and conscientiousness) are based on 6 points Likert scales. There are no significant differences between the groups. There are no significant differences between the two groups.

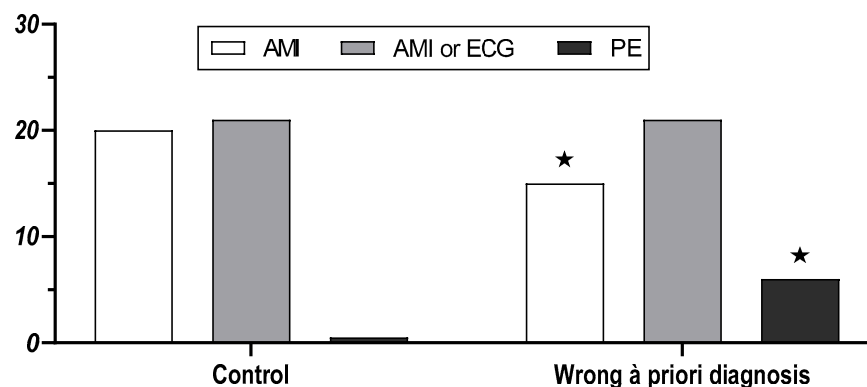
### 3.2. Physicians

The trial was terminated after the interim analysis since the boundary for efficiency, i.e., rejection of the null hypothesis, had been crossed for the primary endpoint (correct diagnosis) and the boundary for futility, i.e., in favour of the null hypothesis, had been crossed for the co-primary endpoint (correct diagnosis or 12-lead ECG). At the time of the interim analysis, 44 (24 female) physicians had been included and randomised to the control group (*n* = 21; 10 female) and to the wrong a priori diagnosis group (*n* = 23; 14 female). All physician participants completed the scenario as intended, no protocol violations occurred, all video-recordings were of sufficient quality for analysis so that data of all 44 participants were analysed and reported. Groups were balanced for personality and years of professional activity after graduation from medical school (10 [IQR 7–16] vs. 8 [IQR 5–17] years; *p* = 0.21). Cohen's kappa for inter-rater reliability was 0.94.

Primary endpoint (Figure 2): A correct final diagnosis was reached by 20/21 (95%) physicians in the control group and 15/23 (65%) in the wrong a priori diagnosis group (*p* = 0.023). In the wrong a priori diagnosis group, 6/23 (26%) indicated pulmonary embolism, i.e., the initially given wrong a priori diagnosis, as their final diagnosis, while 2/23 (9%) indicated another diagnosis. The composite primary endpoint of a correct diagnosis or ordering a 12-lead ECG was reached by 21/21 (100%) in the control group and by 21/23 (91%) in the wrong diagnosis group (*p* = 0.49).

Secondary endpoints: Participants with a correct and final diagnosis did not differ from those with a wrong final diagnosis with regard to gender (*p* = 0.57), personality, and years of professional activity after graduation from medical school (*p* = 0.15). A similar number of questions were asked during history taking in the control group, in the wrong a priori diagnosis group (10 [IQR 8–12] vs. 11 [IQR 8–12]; *p* = 0.64), and from participants with a correct and a wrong final diagnosis in the wrong a priori diagnosis group (11 [IQR 9–13] vs. 10 [IQR 8–12]; *p* = 0.27).





**Figure 2.** Final presumptive diagnosis of the physician participants after a focussed assessment of a patient with AMI in the control group ( $n = 21$ ), and the wrong a priori diagnosis group ( $n = 23$ ). AMI = diagnosis acute myocardial infarction; AMI or ECG = diagnosis acute myocardial infarction OR ordering of a 12-lead ECG; PE = pulmonary embolism. ★ =  $p < 0.05$  vs. both other groups.

In the wrong a priori diagnosis group, only radiation of chest pain (beta 0.65;  $p < 0.0001$ ) reached a  $p$  value of  $< 0.10$  for association with the correct presumptive diagnosis.

Participants were asked three [IQR 1–4] probing questions in the wrong a priori diagnosis group with no significant difference observed between correct or wrong final diagnoses. In the wrong a priori diagnosis group, participants making the wrong final diagnosis did not differ in self-esteem or personality traits from participants making the correct final diagnosis.

Participants took six [IQR 5–7] diagnostic or therapeutic measures with no significant differences between the groups. Like in the student cohort, measures taken by the physicians concurred with their presumptive diagnosis: Participants of the wrong a priori diagnosis group more often monitored oxygen saturation with pulse oximetry (13/23 (57%) vs. 5/21 (24%);  $p = 0.028$ ), measured d-Dimers (12/23 (52%) vs. 3/18 (17%);  $p = 0.009$ ), and examined the patient's legs (8/23 (35%) vs. 1/20 (5%);  $p = 0.016$ ), while participants of the control group administered nitro-glycerine (16/21 (76%) vs. 4/19 (21%);  $p < 0.0001$ ) more often.

#### 4. Discussion

In this prospective randomized trial, a wrong a priori diagnosis resulted in a substantial likelihood of a wrong final diagnosis. In addition, a wrong a priori diagnosis was associated with a different pattern of measures taken during the patient's encounter, including both omissions and commissions. The presence of an a priori diagnosis, regardless whether correct or false, resulted in less questions asked during history taking and probing of a priori diagnoses occurred rarely.

Diagnostic errors can be defined as a diagnosis that is missed, wrong, or delayed, as detected by some subsequent definitive test of finding [22,23]. Most current knowledge on diagnostic errors is based on retrospective and observational studies that are prone to hindsight bias [11]. So far, there are only few prospective data on diagnostic errors and, to the best of our knowledge, this is the first prospective randomized controlled trial on this topic.

In a real-life study, one serious medical error per six to eight patient-days was observed in pediatric intensive care with residents involved in approximately half of the errors [14]. A diagnostic-scenario based prospective study in physicians reported a high rate of incorrect diagnoses (43–56%) [24]. Diagnostic accuracy was positively correlated with the number of critical cue requests and was not affected by the physicians' experience. Of note, approximately 90% of incorrect diagnoses were followed by inappropriate management [24]. A case-vignette-based prospective study in medical students demonstrated a high rate (43%) of diagnostic errors, the main causes of diagnostic errors being inadequate

diagnostic skills, inadequate knowledge, and faulty context generation, while premature closure was noted in 10% [12]. By contrast, the present study demonstrates that knowledge and diagnostic skills did not differ between students making the correct or wrong diagnosis. When faced with unannounced, standardized patients with uncomplicated ambulatory complaints, physicians provided error-free care 73% of the time [15]. However, that percentage decreased to 9% with the addition of bio-medically and contextually complicating factors that required an alternative plan of care. In keeping with our results, diagnostic errors occurred because physicians did not probe in response to contextual red flags [15]. Expert patients' record review followed by a discussion with the treating physician revealed an error rate of 14% and patient harm in 11% of the cases; the most common causes being mistakes, meaning that treating physicians did not realize their actions were incorrect [6].

Diagnostic errors may lead to harm by omission (e.g., delayed or no treatment of the not diagnosed condition) and/or commission (e.g., treatment of the wrongly diagnosed condition) [7]. In the present trial, the wrong a priori diagnosis influenced the measures taken during the triage scenario concurrently to the presumed diagnosis. With the exception of ordering a 12-lead ECG, the pattern of omissions and commissions was very similar in both medical students and physicians. A delayed or initially missed diagnosis in acute myocardial infarction leads to harm since delays in reperfusion therapy results in increased morbidity and mortality [25,26].

History taking and assessment are seen as the most sensitive parts of the diagnostic process [6,27–30]. This process, known as diagnostic reasoning, is a judgement process that takes place under uncertainty and includes considerations of possible diagnoses, harms, and benefits of diagnostic testing, and the patient's preferences and values. So far, diagnostic reasoning is understood as a dual process [31]. It is divided in an analytical deliberate rational process guided by critical thinking and in a non-analytical unconscious, intuitive, and automatic process [27,31,32].

The approach known as pattern recognition, which is a part of the non-analytical system, is a frequently used procedure in the diagnostic process [27,29,33]. Though pattern recognition is a fast and successful method, it can lead to errors due to premature closure [27].

Successful preventive strategies against diagnostic errors require the understanding of underlying mechanisms. There are, however, different kinds of diagnostic errors with different underlying mechanisms [34–39]. The present trial focussed on diagnostic errors induced by or correlated with an a priori diagnosis. Our results are in keeping with the notion that diagnostic errors result less from a lack of medical knowledge than on a misinterpretation of medical facts [40–42]. We observed insufficient probing of a priori diagnosis regardless of whether it was correct or wrong. In addition, differential diagnoses were not systematically assessed regardless of whether there was a correct or wrong a priori diagnosis. Moreover, very few general and unspecific symptoms (e.g., chest pain, dyspnoea) may be perceived as confirmatory proof of an a priori diagnosis. Taken together, an a priori diagnosis is able to induce a fixation error. This fixation error encompasses aspects of a variety of cognitive biases like premature closure, confirmation bias, anchoring bias, and framing effects [38,39]. Asking questions on specific characteristics of chest pain (i.e., quality, radiation) was associated with a high likelihood of making the correct diagnosis. This may be explained by the hypothesis that hearing these typical symptoms from the patient did trigger a pattern recognition process. Surprisingly, the knowledge that the quality of chest pain and the radiation of chest pain are typical symptoms of a myocardial infarction did not translate into asking the corresponding questions during history taking.

Participants in the wrong a priori diagnosis group making the correct or wrong final diagnosis did not differ regarding gender, experience, knowledge, personality traits, quantity of patient's history taking, and several qualitative aspects of history taking like probing. Thus, apparently similar personalities with similar knowledge followed a similar

pattern of assessment and were merely misled or rewarded by the contextual factors like the correctness of their a priori diagnosis or asking few specific questions.

Our findings have several implications: First, as a priori diagnoses are frequent in patients' encounters, practitioners should be aware that in the presence of an a priori diagnosis, diagnostic errors and especially fixation errors may occur easily. Second, preventive measures against diagnostic errors in the presence of an a priori diagnosis are sufficient probing and systematic history taking, which includes asking key questions. This should be emphasized in medical education for both students and physicians. Third, simulation is a suitable tool to expose diagnostic errors and teach and train preventive measures. Fourth, the present trial demonstrates that prospective randomized controlled trials on diagnostic errors are feasible and especially so in the settings of simulation.

Limitations are the single-centre design and the limitations of simulator-based studies, which mainly include the absence of real patients and real environment. However, simulation allowed highly standardised conditions for a large number of participants, which would be very difficult to achieve in real life. Moreover, exposing real patients deliberately to a wrong a priori diagnosis is ethically not justifiable. The inclusion of medical students may be regarded as further limitation. However, the high success rates of the control group, i.e., the group without an a priori diagnosis, demonstrates that our participants, being medical student or physician, had the knowledge and skills to master our triage scenario. The present trial assessed one condition and one wrong a priori diagnosis, and the results can therefore not necessarily be generalized to constellations of other diseases. However, a particular strength of this study is the selection of a scenario (myocardial infarction) and a wrong presumptive diagnosis (pulmonary embolism) that are commonly associated with diagnostic errors in real cases [43].

## 5. Conclusions

The present trial demonstrates that a wrong a priori diagnosis increases the likelihood of a wrong final diagnosis and significantly influences history taking and patient management.

**Author Contributions:** Conceptualization, F.M.L.M., G.M.B., and S.M.; Methodology, G.M.B., K.T., and S.M.; Software, F.M.L.M., M.G.F.; Validation, F.M.L.M., M.G.F. and S.M.; Formal Analysis, F.M.L.M., M.G.F. and S.M.; Investigation, G.M.B., K.T., and S.M.; Resources, S.M.; Data Curation, F.M.L.M., M.G.F. and S.M.; Writing—Original Draft Preparation, F.M.L.M., and S.M.; Writing—Review & Editing, F.M.L.M., G.M.B., K.T., and T.S.; Visualization, F.M.L.M., and S.M.; Supervision, S.M.; Project Administration, S.M.; Funding Acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the regional Ethics Committee (Ethikkommission Nordwest und Zentralschweiz; 85/04; 27.04.04).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data (i.e., video-recordings) are not publicly available due to strict confidentiality guaranteed to participants.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Institute of Medicine, National Academies of Sciences Eam. *Improving Diagnosis in Health Care*; The National Academies Press: Washington, DC, USA, 2015.
2. Newman-Toker, D.E.; Makary, M.A. Measuring diagnostic errors in primary care: The first step on a path forward. Comment on "Types and origins of diagnostic errors in primary care settings". *JAMA Intern. Med.* **2013**, *173*, 425–426. [[CrossRef](#)]
3. Singh, H.; Sittig, D.F. Advancing the science of measurement of diagnostic errors in healthcare: The Safer Dx framework. *BMJ Qual. Saf.* **2015**, *24*, 103. [[CrossRef](#)]

4. Singh, H.; Graber, M.L. Improving Diagnosis in Health Care—The Next Imperative for Patient Safety. *N. Engl. J. Med.* **2015**, *373*, 2493–2495. [[CrossRef](#)]
5. Singh, H.; Giardina, T.D.; Meyer, A.N.; Forjuoh, S.N.; Reis, M.D.; Thomas, E.J. Types and origins of diagnostic errors in primary care settings. *JAMA Intern. Med.* **2013**, *173*, 418–425. [[CrossRef](#)]
6. Zwaan, L.; Thijs, A.; Wagner, C.; Wal, G.; Timmermans, D.R. Relating faults in diagnostic reasoning with diagnostic errors and patient harm. *Acad. Med.* **2012**, *87*, 149–156. [[CrossRef](#)] [[PubMed](#)]
7. Tehrani, A.S.S.; Lee, H.; Mathews, S.C.; Shore, A.; Makary, M.A.; Pronovost, P.J.; Newman-Toker, D.E. 25-Year summary of US malpractice claims for diagnostic errors 1986–2010: An analysis from the National Practitioner Data Bank. *BMJ Qual. Saf.* **2013**, *22*, 672. [[CrossRef](#)]
8. Khoo, E.M.; Lee, W.K.; Sararaks, S.; Abdul Samad, A.; Liew, S.M.; Cheong, A.T.; Ibrahim, M.Y.; Su, S.H.; Mohd Hanafiah, A.N.; Maskon, K. Medical errors in primary care clinics—A cross sectional study. *BMC Fam. Pract.* **2012**, *13*, 1–6. [[CrossRef](#)]
9. Panagioti, M.; Khan, K.; Keers, R.N.; Abuzour, A.; Phipps, D.; Kontopantelis, E.; Bower, P.; Campbell, S.; Haneef, R.; Avery, A.J.; et al. Prevalence, severity, and nature of preventable patient harm across medical care settings: Systematic review and meta-analysis. *BMJ* **2019**, *366*, 14185. [[CrossRef](#)] [[PubMed](#)]
10. Singh, H.; Meyer, A.N.D.; Thomas, E.J. The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving US adult populations. *BMJ Qual. Saf.* **2014**, *23*, 727. [[CrossRef](#)] [[PubMed](#)]
11. Zwaan, L.; Monteiro, S.; Sherbino, J.; Ilgen, J.; Howey, B.; Norman, G. Is bias in the eye of the beholder? A vignette study to assess recognition of cognitive biases in clinical case workups. *BMJ Qual. Saf.* **2017**, *26*, 104. [[CrossRef](#)]
12. Braun, L.T.; Zwaan, L.; Kiesewetter, J.; Fischer, M.R.; Schmidmaier, R. Diagnostic errors by medical students: Results of a prospective qualitative study. *BMC Med. Educ.* **2017**, *17*, 191. [[CrossRef](#)] [[PubMed](#)]
13. Kostopoulou, O.; Delaney, B.C.; Munro, C.W. Diagnostic difficulty and error in primary care—A systematic review. *Fam. Pract.* **2008**, *25*, 400–413. [[CrossRef](#)]
14. Landrigan, C.P.; Rahman, S.A.; Sullivan, J.P.; Vittinghoff, E.; Barger, L.K.; Sanderson, A.L.; Wright, K.P.; O'Brien, C.S.; Qadri, S.; St. Hilaire, M.A.; et al. Effect on Patient Safety of a Resident Physician Schedule without 24-Hour Shifts. *N. Engl. J. Med.* **2020**, *382*, 2514–2523. [[CrossRef](#)]
15. Weiner, S.J.; Schwartz, A.; Weaver, F.; Goldberg, J.; Yudkowsky, R.; Sharma, G.; Binns-Calvey, A.; Preyss, B.; Schapira, M.M.; Persell, S.D.; et al. Contextual errors and failures in individualizing patient care: A multicenter study. *Ann. Intern. Med.* **2010**, *153*, 69–75. [[CrossRef](#)] [[PubMed](#)]
16. Hautz, S.C.; Schuler, L.; Kammer, J.E.; Schaubert, S.K.; Ricklin, M.E.; Sauter, T.C.; Maier, V.; Birrenbach, T.; Exadaktylos, A.; Hautz, W.E. Factors predicting a change in diagnosis in patients hospitalised through the emergency room: A prospective observational study. *BMJ Open* **2016**, *6*, e011585. [[CrossRef](#)]
17. Freund, Y.; Goulet, H.; Leblanc, J.; Bokobza, J.; Ray, P.; Maignan, M.; Guinemer, S.; Truchot, J.; Feral-Pierssens, A.L.; Yordanov, Y.; et al. Effect of Systematic Physician Cross-checking on Reducing Adverse Events in the Emergency Department: The CHARMED Cluster Randomized Trial. *JAMA Intern. Med.* **2018**, *178*, 812–819. [[CrossRef](#)]
18. McCrae, R.R.; John, O.P. An Introduction to the Five-Factor Model and Its Applications. *J. Personal.* **1992**, *60*, 175–215. [[CrossRef](#)]
19. Von Collani, G.; Herzberg, P.Y. Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefuehl von Rosenberg [A revised version of the German adaptation of Rosenberg's Self-Esteem Scale]. *Z. Differ. Diagn. Psychol.* **2003**, *24*, 3–7. [[CrossRef](#)]
20. O'Brien, P.C.; Fleming, T.R. A Multiple Testing Procedure for Clinical Trials. *Biometrics* **1979**, *35*, 549–556. [[CrossRef](#)]
21. Pampallona, S.; Tsiatis, A.A. Group sequential designs for one-sided and two-sided hypothesis testing with provision for early stopping in favor of the null hypothesis. *J. Stat. Plan. Inference* **1994**, *42*, 19–35. [[CrossRef](#)]
22. Graber, M. Diagnostic errors in medicine: A case of neglect. *Jt. Comm. J. Qual. Patient Saf.* **2005**, *31*, 106–113. [[CrossRef](#)]
23. Schiff, G.D.; Kim, S.; Abrams, R.; Cosby, K.; Lambert, B.; Elstein, A.S.; Hasler, S.; Krosnjar, N.; Odwazny, R.; Wisniewski, M.F.; et al. Diagnosing Diagnosis Errors: Lessons from a Multi-institutional Collaborative Project. In *Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology)*; Henriksen, K., Battles, J.B., Marks, E.S., Lewin, D.I., Eds.; Agency for Healthcare Research and Quality (US): Rockville, MD, USA, 2005.
24. Kostopoulou, O.; Oudhoff, J.; Nath, R.; Delaney, B.C.; Munro, C.W.; Harries, C.; Holder, R. Predictors of diagnostic accuracy and safe management in difficult diagnostic problems in family medicine. *Med. Decis. Mak.* **2008**, *28*, 668–680. [[CrossRef](#)]
25. De Luca, G.; Suryapranata, H.; Ottervanger, J.P.; Antman, E.M. Time Delay to Treatment and Mortality in Primary Angioplasty for Acute Myocardial Infarction: Every Minute of Delay Counts. *Circulation* **2004**, *109*, 1223–1225. [[CrossRef](#)]
26. Guerchicoff, A.; Brener, S.J.; Maehara, A.; Witztenbichler, B.; Fahy, M.; Xu, K.; Gersh, B.J.; Mehran, R.; Gibson, C.M.; Stone, G.W. Impact of Delay to Reperfusion on Reperfusion Success, Infarct Size, and Clinical Outcomes in Patients With ST-Segment Elevation Myocardial Infarction: The INFUSE-AMI Trial (INFUSE-Anterior Myocardial Infarction). *JACC Cardiovasc. Interv.* **2014**, *7*, 733–740. [[CrossRef](#)]
27. Coderre, S.; Mandin, H.; Harasym, P.H.; Fick, G.H. Diagnostic reasoning strategies and diagnostic success. *Med. Educ.* **2003**, *37*, 695–703. [[CrossRef](#)] [[PubMed](#)]
28. Gandhi, T.K.; Kachalia, A.; Thomas, E.J.; Puopolo, A.L.; Yoon, C.; Brennan, T.A.; Studdert, D.M. Missed and delayed diagnoses in the ambulatory setting: A study of closed malpractice claims. *Ann. Intern. Med.* **2006**, *145*, 488–496. [[CrossRef](#)] [[PubMed](#)]
29. Goyder, C.R.; Jones, C.H.; Heneghan, C.J.; Thompson, M.J. Missed opportunities for diagnosis: Lessons learned from diagnostic errors in primary care. *Br. J. Gen. Pract.* **2015**, *65*, e838–e844. [[CrossRef](#)]

30. McKinlay, J.B.; Potter, D.A.; Feldman, H.A. Non-medical influences on medical decision-making. *Soc. Sci. Med.* **1996**, *42*, 769–776. [[CrossRef](#)]
31. Lambe, K.A.; Reilly, G.; Kelly, B.D.; Curristan, S. Dual-process cognitive interventions to enhance diagnostic reasoning: A systematic review. *BMJ Qual. Saf.* **2016**, *25*, 808. [[CrossRef](#)]
32. Norman, G.R.; Monteiro, S.D.; Sherbino, J.; Ilgen, J.S.; Schmidt, H.G.; Mamede, S. The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking. *Acad. Med.* **2017**, *92*, 23–30. [[CrossRef](#)]
33. Evans, J.S.; Stanovich, K.E. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspect. Psychol. Sci.* **2013**, *8*, 223–241. [[CrossRef](#)]
34. Crowley, R.S.; Legowski, E.; Medvedeva, O.; Reitmeyer, K.; Tseytlin, E.; Castine, M.; Jukic, D.; Mello-Thoms, C. Automated detection of heuristics and biases among pathologists in a computer-based system. *Adv. Health Sci. Educ. Theory Pract.* **2013**, *18*, 343–363. [[CrossRef](#)]
35. Ely, J.W.; Graber, M.L.; Croskerry, P. Checklists to reduce diagnostic errors. *Acad. Med.* **2011**, *86*, 307–313. [[CrossRef](#)] [[PubMed](#)]
36. Graber, M.L. The incidence of diagnostic error in medicine. *BMJ Qual. Saf.* **2013**, *22*, ii21. [[CrossRef](#)]
37. Ogdie, A.R.; Reilly, J.B.; Pang, W.G.; Keddem, S.; Barg, F.K.; Feldt, J.M.; Myers, J.S. Seen through their eyes: Residents' reflections on the cognitive and contextual components of diagnostic errors in medicine. *Acad. Med.* **2012**, *87*, 1361. [[CrossRef](#)]
38. Saposnik, G.; Redelmeier, D.; Ruff, C.C.; Tobler, P.N. Cognitive biases associated with medical decisions: A systematic review. *BMC Med Inform. Decis. Mak.* **2016**, *16*, 138. [[CrossRef](#)] [[PubMed](#)]
39. Stiegler, M.P.; Ruskin, K.J. Decision-making and safety in anesthesiology. *Curr. Opin. Anaesthesiol.* **2012**, *25*, 724–729. [[CrossRef](#)]
40. Graber, M.L.; Franklin, N.; Gordon, R. Diagnostic error in internal medicine. *Arch. Intern. Med.* **2005**, *165*, 1493–1499. [[CrossRef](#)]
41. Gruver, R.H.; Freis, E.D. A study of diagnostic errors. *Ann. Intern. Med.* **1957**, *47*, 108–120. [[PubMed](#)]
42. Kuhn, G.J. Diagnostic errors. *Acad. Emerg. Med.* **2002**, *9*, 740–750. [[CrossRef](#)]
43. Schiff, G.D. Finding and fixing diagnosis errors: Can triggers help? *BMJ Qual. Saf.* **2012**, *21*, 89–92. [[CrossRef](#)] [[PubMed](#)]