FUNDAÇÃO GETULIO VARGAS
ESCOLA BRASILEIRA DE ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS
MESTRADO EM ADMINISTRAÇÃO

EDUARDO CARDOSO DE MOURA

Rare cases come in pairs! Assessing availability bias in diagnostic reasoning

DISSERTAÇÃO APRESENTADA À ESCOLA BRASILEIRA DE ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS PARA OBTENÇÃO DO TÍTULO DE MESTRE

Rio de Janeiro

2021

EDUARDO CARDOSO DE MOURA

Rare cases come in pairs! Assessing availability bias in diagnostic reasoning

Dissertação apresentada à Escola Brasileira de Administração Pública e de Empresas para a obtenção de grau de mestre.
Orientador: Prof. Dr. Eduardo Bittencourt Andrade

Rio de Janeiro

2021

ACKNOWLEDGMENTS

Two years ago, I decided to take a step towards the master's degree, a decision that might have intrigued many people around me. As a successful entrepreneur and executive, for many people, there was no reason to pursue an academic degree in a business school, especially in social sciences. However, it was a coherent decision for those who really know me as I was looking for personal satisfaction and knowledge, something that somehow was missing in my life.

I already used to make surprising decisions about my career. It was also surprising applying for med school when I was in high school, and quite surprising gradually leaving clinical assistance to become an entrepreneur in a health tech startup. It does not look like a straight and ordinary pathway that brought me here to this point. Still, somehow when I look to the past, I can't imagine being in a better scenario, and getting to this point was only possible because the universe gifted me with the right people in the right places and the right moments—these people who really deserves my acknowledgment.

So, first, I would like to thank my friends and partners in life and business, Pedro Gemal and Bruno Lagoeiro. Watching their growth stimulated me to pursue my own. They always asked me for more, but in the best way, knowing that I was capable of more and giving the best they could to our partnership. Together, we changed healthcare in Brazil, supporting physicians with smarter decisions, a purpose that I will follow for the rest of my life. That was the original inspiration for this work.

Second, I would like to thank those who were presented to me along the way and changed my life, amusing me with this brand new world of behavior and decision-making. My friend Claudio Meilman, who introduced me to this world in the first place, planted this seed in the endless hours of night shift in the intensive care unit. My co-advisor, Professor Rafael Goldszmidt, opened the doors of FGV to me, always supportive and caring. He did much more than I could ask for as an advisor, a true friend to whom I will be eternally grateful. My advisor, Professor Eduardo Andrade, stimulated me to pursue more in turning a research proposal for a discipline in a M.Sc. thesis and a Ph.D. project. And, of course, I can't forget to

thank my colleagues in FGV: Henrique Castro, Jorcianne Ferreira, Manuela Lorenzo, Mariana Furuguem, and Miriam Grobman. We shared more than knowledge; we supported each other in the challenge of remote studying during the pandemic.

Finally, I want to thank those that are the very core of my life, those for whom I live and fight, my family. My lovely wife, Gimelly Barreto, who shared me with the books and articles in the last two years. My sisters, Mariana Moura and Laís Simões, who always care about me and are always there when I need it. My parents, Leonil Moura and Ana Lucia Cardoso, who always believed in education's value and prioritized this value in my life. And my aunts and uncle, Isabel Cardoso, Leila Moura, and Luiz Sérgio Simões, who always supported my education and authentic role models of professionals and academics.

This study is inspired by a popular phenomenon whose description, without any official academic register, is originally attributed to the French surgeon of the 19th century Dr. Alfred Velpeau. Velpeau observed a strange phenomenon regarding rare, odd and singular clinical cases, affirming that they generally occur repeatedly in a short space of time. By affirming that “rare cases come in pairs”, Velpeau created a popular statement that was passed mouth to mouth for hundreds of years, generation after generation, in hospital halls. Velpeau was not aware, however, that he and all the following generations of doctors were deceived by a cognitive bias: the availability bias.

RESUMO

Erros de diagnóstico são um problema comum e caro na assistência em saúde. A literatura demonstra de forma consistente que vieses cognitivos podem estar presentes em muitos casos de erros diagnóstico em medicina. Um dos vieses mais comuns na prática clínica é o viés de disponibilidade, a tendência de pesar a probabilidade de um evento pela facilidade com que as instâncias vêm à mente. Para avaliar esse viés na prática clínica, propomos uma nova abordagem experimental: selecionar médicos com base em suas experiências recentes tratando pacientes com algumas doenças de interesse, e examinando seu desempenho diagnóstico em um novo caso clínico semelhante ao que assistiram na vida real, porém com um diagnóstico mais provável diferente. Também propomos um exercício de avaliação da frequência das doenças, realizado antes do diagnóstico do caso clínico, seguindo a hipótese de que este tipo de intervenção de raciocínio deliberado pode suprimir o efeito do viés. Verificamos que o viés de disponibilidade aumentou a probabilidade de escolha da doença saliente como diagnóstico do caso, mas não afetou a probabilidade de escolha do diagnóstico correto. Não encontramos suporte para nossa intervenção na avaliação da frequência das doenças para neutralizar esse viés. Também observamos que médicos experientes são mais propensos ao viés de disponibilidade do que médicos novatos.

Palavras-chave: viés de disponibilidade, heurística de disponibilidade, tomada de decisão clínica, erros de diagnóstico, raciocínio diagnóstico.

ABSTRACT

Diagnostic errors are a common and costly issue in healthcare practice. Previous literature consistently demonstrates that cognitive biases may be present in a large number of cases of medical diagnostic errors. One of the most common biases in clinical practice is the availability bias, or the tendency of weighing the likelihood of an event by the ease with which instances come to mind. To assess this bias in clinical practice, we propose a novel experimental approach: selecting physicians based on their recent experiences treating patients with some diseases of interest and examining their diagnostic performance in a new clinical case similar to the one they experienced in real life but with a different most probable diagnosis. We also propose a base rate exercise, performed before the clinical case diagnosis, hypothesizing that this kind of deliberate reasoning intervention might suppress the influence of the availability bias. Our results show that availability bias increased the likelihood of choosing the salient disease as the diagnostic of the case, but it didn´t affect the probability of choosing the most likely diagnosis. We found no support for our base rate intervention in counteracting this bias. We also observed that experienced physicians were more prone to bias than novice physicians.

Keywords: availability bias, availability heuristic, clinical decision-making, diagnostic errors, diagnostic reasoning.

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pelo Sistema de Bibliotecas/FGV

Moura, Eduardo Cardoso de
Rare cases come in pairs! Assessing availability bias in diagnostic reasoning / Eduardo Cardoso de Moura. – 2021.

51 f.

Dissertação (mestrado) - Escola Brasileira de Administração Pública e de Empresas, Centro de Formação Acadêmica e Pesquisa.
Orientador: Eduardo Bittencourt Andrade.
Inclui bibliografia.

1. Erros de diagnóstico. 2. Erros de diagnóstico - Processo decisório. 3. Clínica médica - Estudo de casos. I. Andrade, Eduardo Bittencourt. II. Escola Brasileira de Administração Pública e de Empresas. Centro de Formação Acadêmica e Pesquisa. III. Título.

CDD – 610.696

Elaborada por Maria do Socorro Almeida – CRB-7/4254

FUNDAÇÃO GETULIO VARGAS
MESTRADO EM ADMINISTRAÇÃO
ESCOLA BRASILEIRA DE ADMINISTRAÇÃO PÚBLICA E DE EMPRESAS

EDUARDO CARDOSO DE MOURA

"RARE CASES COME IN PAIRS! ASSESSING AVAILABILITY BIAS IN DIAGNOSTIC REASONING".

DISSERTAÇÃO APRESENTADO(A) AO CURSO DE MESTRADO EM ADMINISTRAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE(A) EM ADMINISTRAÇÃO.

DATA DA DEFESA: 08/06/2021

ASSINATURA DOS MEMBROS DA BANCA EXAMINADORA

PRESIDENTE DA COMISSÃO EXAMINADORA: PROFº/ª EDUARDO BITTENCOURT ANDRADE

PROFº/ª EDUARDO BITTENCOURT ANDRADE
ORIENTADOR(A)

PROFº/ª RAFAEL GUILHERME BURSTEIN GOLDSZMIDT
MEMBRO INTERNO

PROFº/ª SÍLVIA MAMEDE STUDART SOARES
MEMBRO EXTERNO

RIO DE JANEIRO, 08 DE JUNHO DE 2021.

PROFº/ª FLAVIO CARVALHO DE VASCONCELOS
DIRETOR(A)

PROFº ANTONIO DE ARAUJO FREITAS JUNIOR
PRÓ-REITOR DE ENSINO, PESQUISA E PÓS-GRADUAÇÃO

CONTENTS

1. INTRODUCTION

Clinical decision-making is the process of analyzing and comparing the options, possibilities, uncertainties, and risks for choosing a proper course of action (Croskerry, 2003). Although humans used to be considered as fully informed and rational actors by traditional economics theory, the advent of behavioral economics, conversely, showed us that this is not quite accurate, and decision-making is susceptible to common biases that contradict the vision of the rational actor (Ariely, 2008).

As exemplars of human rationality, clinicians are also prone to engage in automatic behaviors, being influenced by cognitive, emotional, and social factors (Blumenthal-Barby & Krieger, 2015) that mislead them away from the optimal decision. Behavioral economics, however, also suggests that those biases are predictable (Thaler & Sunstein, 2008), which allows us to anticipate and intervene to counteract them.

Previous evidence shows that cognitive factors are involved in around three-quarters of diagnostic errors in clinical practice (Graber et al., 2005; Singh et al., 2013). Although huge effort has been made on research of patient safety interventions since the release of *To Err is Human* (Kohn et al., 1999), less attention has been made to cognitive issues (Newman-Toker & Provonost, 2009).

Diagnostic errors are a common and costly issue in healthcare practice. Berner and Graber's (2008) review of diagnostic errors found that the literature consistently demonstrates a rate up to 10-15% of diagnostic errors in most medical fields, and diagnostic errors remain the most common type of error in medical malpractice (Weingart, 2000; Graber et al., 2002; Graber, 2013).

Diagnostic reasoning follows a 'dual-process' model of thinking that has been extensively studied in psychology in the last 30 years, and recently developed into a theoretical model to understand physician's performance in diagnosis (Croskerry, 2009).

The dual-process theory of diagnostic reasoning assumes the existence of two parallel systems: System 1, characterized by a non-analytical approach that uses heuristics as

shortcuts, rapid, unconscious, and contextualized, for judgments; and System 2, characterized by an analytical reasoning, slow, deliberative, logical, and conceptual (Evans, 2008).

Most of the time, System 1 processing is quite useful as it demands low effort and allows fast responsiveness. For instance, most physicians promptly recognize, through pattern recognition, a patient with the combination of signs and symptoms of an acute myocardial infarction. Nevertheless, heuristics occasionally lead to serious and systematic errors (Tversky & Kahneman, 1974), known as bias, when, in the same context, an acute and potentially deadly disease, like for instance aortic dissection, is hidden behind a similar presentation. In this case, physicians would rely on deliberate reasoning to override System 1 processing, detect the correct diagnosis, and avoid the biased judgment.

Previous literature recognizes cognitive biases as a common source of error in reasoning processes. More than 40 types of biases that may affect clinical reasoning have been already described (Croskerry, 2003; Blumenthal-Barby & Krieger, 2015), and three of them seem particularly prevalent and powerful: anchoring bias, availability bias, and framing effects (Ogdie et al., 2012). Not just these biases are proven to be very common as they are also proven to cause diagnostic inaccuracy. A systematic review found that cognitive biases (overconfidence, lower tolerance to risk, the anchoring effect, and availability bias) were associated with diagnostic inaccuracies in 36.5 to 77 % of the cases presented (Saposnik et al., 2016).

Due to its importance, the current study focuses on the availability bias, the resulting bias from the misuse of the availability heuristic, which is defined as the tendency of weighing the likelihood of the occurrence of an event "by the ease with which instances or associations come to mind." (Tversky & Kahneman, 1973).

In clinical context, by using the availability heuristic, the likelihood of a diagnosis might be decided based on how easily examples of that diagnosis come to mind, what may eventually lead to biased decisions. For example, a physician that recently diagnosed an aortic dissection, a relatively uncommon and menacing disease, may have this diagnosis more salient, and, consequently, consider that a new case of chest pain is aortic dissection, despite acute coronary artery disease being a much more probable diagnosis.

We propose a novel experimental approach to assessing availability bias in diagnostic reasoning in clinical practice, and, this way, aiming on making the following contributions: (1) examining how clinical assistance in the real world induce availability bias; (2) counteracting the bias with a base rate pre-exercise; and (3) examining how physician's clinical experience interacts with the bias. In the next section, first, we will discuss "state-of-the-art" evidence regarding availability bias in diagnostic reasoning. Then, we identify current literature gaps, upon which we build our contributions.

2. THEORETICAL BACKGROUND

2.1. Diagnostic Reasoning

Using a decision theory perspective, diagnostic reasoning is the process of updating opinion when new imperfect information - known as clinical evidence - appears. Using Bayes' theorem for this task, the pretest probability is the diagnostic probability of a disease before new information is acquired, which usually corresponds to the known prevalence/incidence of the disease or the physician's subjective impression of it. And the post-test probability is a function of the pre-test probability and the strength of the new evidence, or, in other words, the diagnostic probability of a disease given the new clinical evidence (Elstein & Schwarz, 2002).

Thus, using Bayes' theorem, it is possible to identify two major classes of errors in diagnostic reasoning: assessing pretest probability and assessing the strength of the evidence. Our approach on this study is to examine if and how availability bias is applied as a simplifying rule (heuristic) on judging a clinical case diagnostic probability, influencing the pre-test probability and the strength of the evidence, and, consequently, the post-test probability.

As previously mentioned, diagnostic reasoning follows a dual-process model (Croskerry, 2009). System 1 uses an intuitive approach, leaning heavily on the experience of the decision maker in pattern recognition. Thus, it is likely that experienced physicians use more frequently pattern recognition, reserving the analytical approach (System 2, known as hypothetico-deductive) only with difficult cases. Therefore, viewing the process of diagnostic reasoning as pattern recognition is considering it a process of categorization, in which a new case is categorized by its resemblance to memories of instances previously seen or to a generic prototypical exemplar of the disease (Brooks et al., 1991).

2.2. Availability heuristic

In clinical context, the availability heuristic – that is, the tendency of weighing the likelihood of occurrence of an event by easiness of recalling instances – is often helpful, because clinical situations that come to mind easily generally occur more frequently. However, retrievability is

not only related to frequency, which in this case would give us the correct estimate of pre-test probability, but it is also often related to familiarity, saliency (where one instance elicits more attention than another), and recency (Dimara et al., 2014).

Therefore, availability heuristic may increase the judgment of likelihood of a particular disease by the ease by which one reminds of recent experiences with similar cases (pattern recognition), ignoring the real base rate for the diagnostic hypothesis (pre-test probability), and eventually leading to a biased decision (wrong diagnosis). Under normal conditions, this non-analytical heuristic-based judgement of System 1 is suppressed by the analytical approach of System 2 (Croskerry, 2009). But under conditions of time constraints and fast decision-making, in the absence of System 2 deliberate process, this heuristic may lead to availability bias.

The literature has so far demonstrated the effect of availability heuristic by manipulating physicians' exposure to a particular disease in lab experiments and measuring diagnostic accuracy next. For example, in an experimental approach, Mamede et al. (2010) induced availability bias in physicians using vignettes of clinical cases and, next, measured their diagnostic accuracy in similar clinical cases but with distinct correct diagnosis. Consistent with the availability bias, they found that second-year residents scored lower on the cases similar to the previously encountered than on the other cases. Interestingly, they did not found evidence among first-year residents, a result also indicating a possible effect of experience. Finally, in their study, they were also able to show the effect of deliberate reasoning on increasing diagnostic accuracy by inducing reflective reasoning in a third phase (post-diagnosis).

Another related work is the study of Schmidt et al. (2014), who also found a substantial availability bias effect induced not by exposure to a clinical case in a vignette but by exposure to media-provided information about a disease as a cause of diagnostic errors in a lab experiment. The authors also counteracted the bias by a reflection activity, revealing how availability bias is apparently associated with non-analytical reasoning and proving that it can be suppressed by the activation of System 2.

Besides counteracting it with deliberate reasoning, metacognitive interventions were also proposed and tested. Interventions succeeded in increasing physicians' awareness of biases

indeed (Bond et al., 2004; Reilly et al., 2013), although they failed in improving diagnostic accuracy (Sherbino et al., 2014; Norman et al., 2017; Lambe et al., 2016).

In a new approach on preventing bias, Mamede et al. (2020) conducted an "immunization" intervention directed at refining physicians' knowledge of how to discriminate between similar diseases 1-week before a diagnostic test. Knowledge of clinical features that discriminate between similar-looking diseases predicted susceptibility to bias among physicians with similar clinical experience. Physicians with less knowledge of discriminating features (control group) performed poorly in a diagnostic test (using the same clinical vignettes from Mamede et al. 2010) due to availability bias (58% more errors) in comparison to the treatment group (who took a course on distinctive features of the diseases 1-week before).

2.3.Hypothesis and contributions

Despite these advancements, many questions remain unanswered. First, availability bias has been manipulated under lab experiments, specially using clinical vignettes, but, so far, there is no empirical evidence that real cases may play the same role on inducing bias.

In this work, aiming at filling this gap, we assess availability bias induced by the real experience of the physician in clinical practice, by using a secondary dataset to select physicians that were recently exposed to patients with a particular disease and thus are prone to availability bias (bias-prone condition) and comparing their diagnostic accuracy to physicians that did not attend patients with the particular disease (control condition). Note that relying on physicians' personal exposure to patients with a given disease is important because, theoretically, opposite forces may be at play. Assuming those in the bias-prone condition have also learned more about the target disease, they may well become more knowledgeable and, hence, more (rather than less) accurate in future diagnosis where the target disease is present. Put simply, in theory, increased knowledge may mitigate or even reverse the impact of saliency, and the subsequent availability bias. In any case, following the bulk of the literature, we expect that the recent experience of attending a patient with the disease to increase retrievability due to recency and saliency of the disease (bias-prone

condition), and bias judgment accordingly despite the potential opposing influence of increased knowledge. Therefore, we hypothesize that:

Hypothesis 1: Inaccurate diagnosis will be higher in the bias-prone condition (vs. control condition), driven at least in part by a relatively higher number of physicians indicating that the salient disease rather than the most probable disease is the accurate diagnosis.

Second, researchers proposed that reflective reasoning mitigate the effect of availability bias after a bias-prone decision has already been made, and then physicians reconsider their initial diagnosis (Mamede et al., 2010; Schmidt et al., 2014). Interventions on preventing bias, however, have found limited results (Bond et al., 2004; Reilly et al., 2013; Norman et al., 2017; Lambe et al., 2016; Sherbino et al., 2014). Therefore, considering the known mechanism of availability bias of "easiness of recall" (Schwarz & Vaughn, 2002) as being an attribute substitution to the base rate consideration for a particular disease, we propose that an exercise inducing deliberate reasoning before the actual diagnostic decision may prevent bias from happening as it makes promptly available the base rate of the disease, an information that plays against System 1 associative availability heuristic. To induce deliberate reasoning, we asked participants in the base rate condition to evaluate how frequent is the incidence of the diseases under evaluation prior to the diagnosis assessment, in the hope that base rate knowledge would suppress bias (System 2 suppressing System 1). Thus, in hypothesis 2, we hypothesize that:

Hypothesis 2: The main effects predicted in Hypothesis 1 will be reduced when participants are asked (vs. not asked) to estimate the base rate for each of the possible diseases prior to diagnosis assessment.

In lab experiments performed so far, by inducing availability bias through clinical vignettes or media-provided information, the effect is examined immediately or only a few hours later after the bias-inducing event (Mamede et al., 2010; Schmidt et al., 2014; Mamede et al., 2020). Thus, it is also not clear for how long after the exposure to the clinical case physicians remain vulnerable to bias. In our sample, we select physicians for the bias-prone condition who had attended patients suspected or confirmed with the disease in the last 7 days. Although much larger than the usual time frame used in previous experiments, we believe it is a much more reasonable time frame to check for availability in clinical practice, as physicians don't regularly receive similar patients every hour.

Going further on the understanding of physicians' susceptibility to availability bias, it is also not completely clear whether and how clinical experience influences propensity to bias. Mamede et al. (2010) found greater propensity to availability bias in second-year residents in comparison to first-year residents. This result is consistent with theoretical work on availability heuristic posing that "people rely on their subjective recall experiences only when they consider themselves knowledgeable" (Schwarz & Vaughn, 2002). Thus, it is reasonable to accept that reliance on non-analytical reasoning tends to increase with experience (Schmidt et al., 1990). However, to the best of my knowledge there are no empirical studies comparing physicians with many years of clinical practice with residents or recently graduated physicians.

In summary, this study aims at making the following contributions to literature: (1) We examine if and how clinical experiences in the real world induce availability bias in diagnostic reasoning in the judgement of the most probable diagnosis in simulated clinical cases using vignettes; (2) we examine the effect of a compulsory base rate evaluation pre-test exercise in preventing availability bias and increasing diagnostic accuracy; (3) in an exploratory approach, we evaluate the interaction of clinical experience with the availability bias, examining if more experienced physicians are more prone to bias than their less experienced counterparts.

3. METHODS

In this study, we perform a survey-experiment to assess availability bias induced by recent clinical cases attended by physicians. We perform a 2 (bias-prone condition vs. control) x 2 (base rate assessment vs. no assessment) between-subjects design with four replicates (clinical cases). The study was approved by FGV ethics' committee (n. 217/2020) and was pre-registered at *Aspredicted.com* (https://aspredicted.org/blind.php?x=pi24tu).

3.1.Participants

Participants were selected for invitation based on their recent access to a digital platform of medical information (Whitebook Clinical Decision®). Physicians consult this digital platform in order to solve clinical questions regarding diagnosis and treatment of diseases in clinical practice. Consultation of the platform is a reliable proxy of what physicians are facing in their clinical practice, as content resources are specifically designed to be consulted in point-of-care and are much less often consulted for study purposes unrelated to a specific patient. Thus, we used the access to the system to identify physicians that were likely to have recently attended a patient with a possible diagnosis of a given disease.

Physicians who had accessed one and only one of the following four diseases of interest (*Acute pancreatitis, Diabetic ketoacidosis, Acute appendicitis, and Herpes zoster*) in the medical information platform in the last 5 days were invited to participate. These diseases were chosen due to their relatively high frequency in clinical practice, and consequently being highly demanded for consultation in the medical content platform.

Although one might hypothesize that rare diseases might elicit more salience when they show up in clinical practice, if we had chosen rarer diseases, it would be harder to obtain the necessary sample size. By choosing relatively common diseases, we also opted by a much more conservative scenario, as common diseases are less salient to the physician when they show up in clinical practice. Thus, by finding that even relatively common diseases are able to become relatively salient and, consequently, elicit availability bias, we can make a claim that rarity is not a necessary condition. To guarantee the need participation in the control

condition, we selected as a second group the same number of physicians using similar criteria, however making sure that they had not accessed content of any of these diseases under study.

3.2. Sample size

To identify our necessary sample size, we performed an exploratory pre-test, in which the proportion of incorrect diagnosis in the bias-prone condition was 15% higher than in the control condition. Power calculations indicated a sample size of 136 participants on each condition to achieve 0.80 power to detect a significant effect of this magnitude (alpha = .05). Concerning the interaction proposed in hypothesis 2, to identify an interaction that eliminates the effect of bias in the base-rate assessment condition, a sample of 325 participants in each one of the four conditions would be necessary, which defines our sample size of 1300 participants. In order to achieve this number of participants, participation in the survey-experiment was offered on a weekly-basis to all physicians that accessed only one of the four diseases of interest in the last 5 days (bias-prone condition), and to a random group of physicians (with the same sample size) that did not access any of the diseases, until we reach the total of 1300 clinical cases solved.

3.3. Survey-experiment

In the survey-experiment, each participant was presented with one hypothetical clinical case in the form of a vignette. For the clinical case, the physician had to indicate - out of four options - the most probable diagnosis, by answering the question "What is the most probable diagnosis for this case?". For each case, there were four options presented in random order: (a) the most common disease, and (b) the salient disease in the bias-prone condition, and two other possible diagnoses used as fillers. All clinical cases were created in consultation with a board of four internal medicine professors for approval. Clinical vignettes were presented in Portuguese. Portuguese and English translations are attached in Appendix B.

Observation of Bias Prone vs. Control conditions. Once physicians accepted the invitation by email, available up to 48 hours after the email was sent, they were directed to an online survey. In the survey, the physician visualized one of four clinical cases in the form of a

vignette. Each case had, as one of the alternative diagnostics, one and only one of the four diseases considered in participants selection (*Acute pancreatitis, Diabetic ketoacidosis, Acute appendicitis, and Herpes-zoster*). Physicians who had accessed one of these four diseases of interest were purposely assigned to a case which included this disease as one of the alternative diagnostic options, but with a different most probable diagnosis. Physicians selected for not having accessed any of the diseases of interest were randomly assigned to one of the same four cases.

While consulting the system was used to identify physicians likely to have attended a patient suspect of a given disease, the operationalization of the bias-prone condition (vs control) was based on the answer to a question at the end of the survey asking if the physician indeed attended a patient with the disease of interest. Thus, only participants that confirmed having attended a patient (regardless of having consulted the disease in the system, including those in the control group that by chance attended such a patient) were assigned to the bias-prone condition. Our assignment process is summarized in Appendix A.

Base rate manipulation. Participants were randomly assigned to two different conditions, with (vs. without) a base rate assessment before the clinical case vignette. The base rate assessment was a question of "On a scale of 1 to 5 where "1 = Very rare" and "5 = Very common", check how frequent are the following diseases?" followed by the 4 diseases corresponding to the options of the clinical case in a random order, to be evaluated in a 5-point Likert scale.

After the base rate assessment (when applicable) and the clinical case vignette diagnosis, participants answered demographic questions (age and gender), and professional questions (year of completion of medical school, degree of medical formation, and specialty). Finally, by the end of the survey, participants were asked to answer the question "Have you attended any patient with "*X*" in the last 7 days?", where *X* is the salient disease under consideration, with 3 different possible responses: "No, I did not attend", "I attended a patient suspect of the disease not confirmed to the present moment", "Yes, I attended a patient with confirmed diagnosis of the disease". Participants were considered as having recent real experiences, and thus assigned to the bias-prone condition, if they had attended patients with suspected or confirmed diagnosis of the disease, otherwise they were assigned to control condition.

Once the respondents performed a step, like the base rate assessment (when applicable) or the clinical case vignette diagnosis, they could not go back and review their responses.

As an incentive for participating in the survey-experiment, all respondents received, by the end of the survey, a coupon for 30-day free access to the medical information digital platform.

3.4.Data analysis

Our dependent variables were binary, indicating if (a) the most-likely diagnosis was chosen or not, and (b) if the salient disease was chosen or not. Our independent variables of interest were also binary indicating: (a) if the participant was in the bias-prone condition or in the control condition; and (b) if the participant was assigned to the base rate assessment or not. As controls we considered dichotomous variables representing the clinical cases, and a dichotomous variable indicating if the respondent is a specialist in the medical area related to the vignette or not. We used chi-square tests and logistic regression models to test the hypotheses.

To test hypothesis 1, we performed chi-square tests and logistic regressions to assess the main effect of availability bias towards choosing (a) the most-likely diagnosis and (b) the salient disease. In the logistic regression, we included as a predictor a binary variable indicating if the participant was in the bias-prone condition (or in the control condition), and controls for the case type (with a set of dummy variables representing the different clinical cases) as well as for the specialty of the respondent (with a dummy variable indicating if the respondent is a specialist in the medical area related to the vignette or not).

And to test hypothesis 2, we included in the models a binary variable indicating if participant performed the base rate assessment or not, the bias-prone condition variable and their interaction term.

Additionally, we performed an exploratory analysis of the moderating role of clinical experience on the effect of availability bias in our dependent variables, by clustering physicians with up to 5 years since graduation (novice) and comparing to physicians with more than 5 years since graduation (experienced), including this binary variable and the interaction of this variable with bias-prone condition variable in the model. In Brazil, the first 5 years of clinical practice, on average, are related to physician's specialization, and that is the reason why we chose 5 years as our cut-off of novices and experienced physicians.

4. RESULTS

Our survey was responded by 1,302 participants. However, 10 participants were not physicians and for this reason were removed from our final sample. Therefore, we proceeded our analysis using the final sample of 1,292 physicians.

Our selection process, as described in the methods section, involved inviting physicians who recently consulted a digital platform of medical content in 2 different ways: (1) Physicians who had accessed one of the four diseases of interest and thus, in this group, we expected most of the respondents had attended the disease of interest in the last 7 days (bias-prone condition); and (2) Physicians who had accessed contents other than the four diseases of interest, expecting that most of the respondents had not attended the disease of interest in the last 7 days (control condition). As expected, in the first group, 68% of the physicians had attended a patient with the disease of interest suspected or confirmed; while in the latter only 20% had attended by chance a patient with the disease of interest suspected or confirmed. Therefore, our selection process was effective in obtaining respondents in bias-prone condition and control condition.

4.1.Descriptive statistics

To assess the comparability of the two groups of physicians, we performed a descriptive comparison in terms of age, gender, and medical degree between those in bias-prone and those in control conditions. Descriptive comparison is reported next in table 1.

We also performed a balance check comparing the participants randomly assigned to different base rate assessment conditions. Descriptive comparisons are also reported in Table 1.

Table 1 – Descriptive statistics by conditions

	Bias-prone condition (N = 563)	Control condition (N = 729)	Standard. difference*	Base rate assessment (N = 669)	No assessment (N = 623)	Standard. difference*
Age, mean (SD), y	31.4	32.4	-0.11	31.6	32.4	-0.09
Gender, No (%)						
Women	300 (53.3)	364 (49.9)	0.08	351 (52.5)	313 (50.2)	0.04
Men	261 (46.4)	362 (49.7)	-0.08	313 (46.8)	310 (49.8)	-0.05
Other/Not answered	2 (0.4)	3 (0.4)	0.01	5 (0.8)	0 (0.0)	0.11
Medical degree, No. (%)						
Generalists	346 (61.5)	403 (55.3)	0.14	409 (61.1)	340 (54.6)	0.14
Currently specializ.	96 (17.1)	140 (19.2)	-0.08	120 (17.9)	116 (18.6)	-0.02
Specialists	121 (21.5)	186 (25.5)	-0.09	140 (20.9)	167 (26.8)	-0.15

* difference between means (or proportions) divided by the square root of the weighted average of the within group variance.

Comparing the observed bias-prone condition with control condition (Table 1), we found very small differences of age, gender, and medical degree (all differences smaller than .15 standard deviations). Comparing the randomized conditions (base rate assessment vs. no assessment), we also found differences of small magnitude (all under .15 standard deviations).

Next, in table 2, we present the summary statistics and correlations between the variables used in our models.

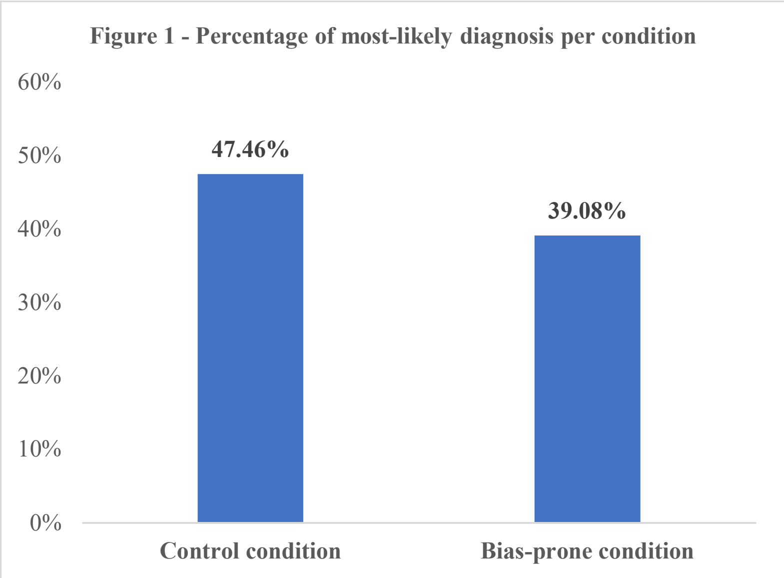
Table 2 – Summary statistics and correlation matrix

Variables	Mean (SD)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Correct diagnosis	0.44 (0.50)							
(2) Salient disease diagnosis	0.40 (0.49)	-0.73*						
(3) Bias-prone condition dummy	0.44 (0.50)	-0.08*	0.10*					
(4) Base rate assessment dummy	0.52 (0.50)	0.02	-0.06*	0.06*				
(5) Specialties match	0.03 (0.17)	0.09*	-0.06*	0.08*	-0.02			
(6) Experience > 5 years dummy	0.22 (0.42)	0.04	0.01	-0.03	-0.02	0.00		
(7) Time since system access	2.76 (1.63)	-0.03	0.01	0.11*	0.02	0.05	0.02	
(8) Age	32.0 (8.35)	0.01	0.01	-0.06*	-0.05	-0.01	0.60*	0.00

* $p<0.05$

4.2.Assessing diagnostic accuracy (H1)

In order to assess the effect of bias-prone condition on the correct diagnosis, we compared the proportions of correct diagnosis between the bias-prone condition and control condition using a chi-square test. The results show that the correct diagnosis was dependent on condition ($\chi2$ (1) = 9.08, p < .01), with a higher percentage of correct diagnosis given by physicians in control condition (47.46%) than by physicians in bias-prone condition (39.08%, Fig. 1).



This result was confirmed in a logistic regression including, as a predictor, a binary variable indicating if the participant was in the bias-prone condition [1 = yes; 0 = no], and, as the dependent variable, a dummy indicating if the participant chose the most-likely diagnosis [1 = yes; 0 = no], (β = -0.34, SE = 0.11, p < .01, Table 3, model 1A). With the addition of controls - a binary variable for each clinical case vignette and a binary variable for the match of the specialty of the respondent with the specialty of the clinical case [1 = same specialty; 0 = different specialties] – the coefficient remains in the same direction but loses statistical significance (β = -0.17, SE = 0.14, p = 0.23, Table 3, model 1B). It is also noteworthy that the variables representing the cases have significant and large coefficients, indicating that the likelihood of finding the correct diagnosis varied a lot across cases (with case 3 – bias-prone condition: *Herpes zoster* – and 4 – bias-prone condition: *Acute appendicitis* – having lower rates compared to case 1 – bias-prone condition: *Acute pancreatitis* – and 2 – bias-prone condition: *Diabetic ketoacidosis*).

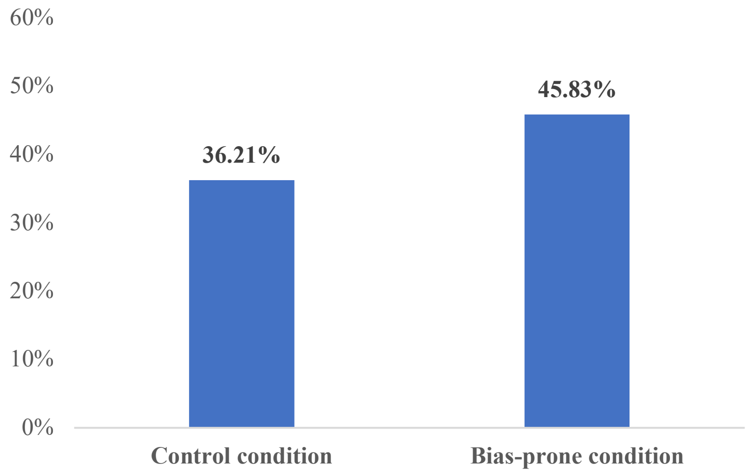
Table 3 – Logistic regression models for most-likely diagnosis (DV)

	DV: Most-likely diagnosis					
	Model 1A	**Model 1B**	**Model 1C**	**Model 1D**	**Model 1E**	**Model 1F**
Bias-prone condition (vs. Control)	-0.34**	-0.17	-0.35**	-0.17	-0.24	0.05
	(0.11)	(0.14)	(0.11)	(0.14)	(0.17)	(0.20)
Base rate assessment (vs. No assessment)			0.08	0.03	0.17	0.21
			(0.11)	(0.14)	(0.15)	(0.18)
Bias-prone*Base rate					-0.20	-0.42
					(0.23)	(0.28)
Specialties match		0.94*		0.94*		0.94*
		(0.41)		(0.41)		(0.40)
Clinical cases						
Case 2		0.29		0.29		0.29
		(0.18)		(0.18)		(0.18)
Case 3		-2.73***		-2.73***		-2.74***
		(0.19)		(0.19)		(0.20)
Case 4		-0.88***		-0.88***		-0.88***
		(0.20)		(0.20)		(0.20)
N	1292	1292	1292	1292	1292	1292
Pseudo R-squared	0.01	0.25	0.01	0.25	0.01	0.25

Standard errors between parentheses
**** p < .001, ** p < .01, * p < .05, † p < 0.10*

Regarding the effect of availability bias in choosing the salient disease as the probable diagnosis, we compared the proportions of salient disease diagnosis between the bias-prone condition and control condition using a chi-square test. The results show that salient disease diagnosis was dependent on condition ($\chi 2$ (1) = 12.19, p < .001), with a lower percentage of salient disease diagnosis given by physicians in control condition (36.21%) than by physicians in bias-prone condition (45.83%, Fig. 2).

Figure 2 - Percentage of salient disease diagnosis per condition



This result was confirmed by a logistic regression including the choice of the salient disease as the dependent variable and a binary variable indicating the bias-prone condition as the predictor (β = 0.40, SE = 0.11, p < .001, Table 4, model 2A) and also by a model including the aforementioned controls (β = 0.29, SE = 0.12, p = < .05, Table 4, model 2B). Once again, there was significant heterogeneity across cases, with case 3 (bias-prone condition: *Herpes zoster*) presenting a much higher rate of choice of the salient disease.

Thus, we found empirical evidence to partially support H1 as we found significant results for the choice of the salient disease, but not regarding the choice of the most-likely diagnosis in the model with control variables.

Table 4 – Logistic regression models for salient disease diagnosis (DV)

	DV: Salient disease diagnosis					
	Model 2A	Model 2B	Model 2C	Model 2D	Model 2E	Model 2F
Bias-prone condition (vs. Control)	0.40***	0.29*	0.42***	0.30*	0.42*	0.27
	(0.11)	(0.12)	(0.12)	(0.12)	(0.17)	(0.18)
Base rate assessment (vs. No assessment)			-0.26*	-0.26*	-0.26†	-0.29†
			(0.11)	(0.12)	(0.16)	(0.17)
Bias-prone*Base rate					0.00	0.07
					(0.23)	(0.25)
Specialties match		-0.67†		-0.69†		-0.69†
		(0.41)		(0.41)		(0.41)
Clinical cases						
Case 2		-0.33†		-0.32		-0.32
		(0.20)		(0.20)		(0.20)
Case 3		1.47***		1.46***		1.47***
		(0.17)		(0.17)		(0.17)
Case 4		0.33		0.32		0.32
		(0.21)		(0.21)		(0.21)
N	1292	1292	1292	1292	1292	1292
Pseudo R-squared	0.01	0.11	0.01	0.11	0.01	0.11

Standard errors between parentheses

**** p < .001, ** p < .01, * p < .05, † p < 0.10*

4.3.Assessing base rate intervention (H2)

Regarding the effect of the base rate assessment in reducing the availability bias, we observed the proportion of (a) correct diagnosis and (b) salient disease diagnosis reported in Fig. 3 and 4, respectively.

Figure 3 - Percentage of most-likely diagnosis per group

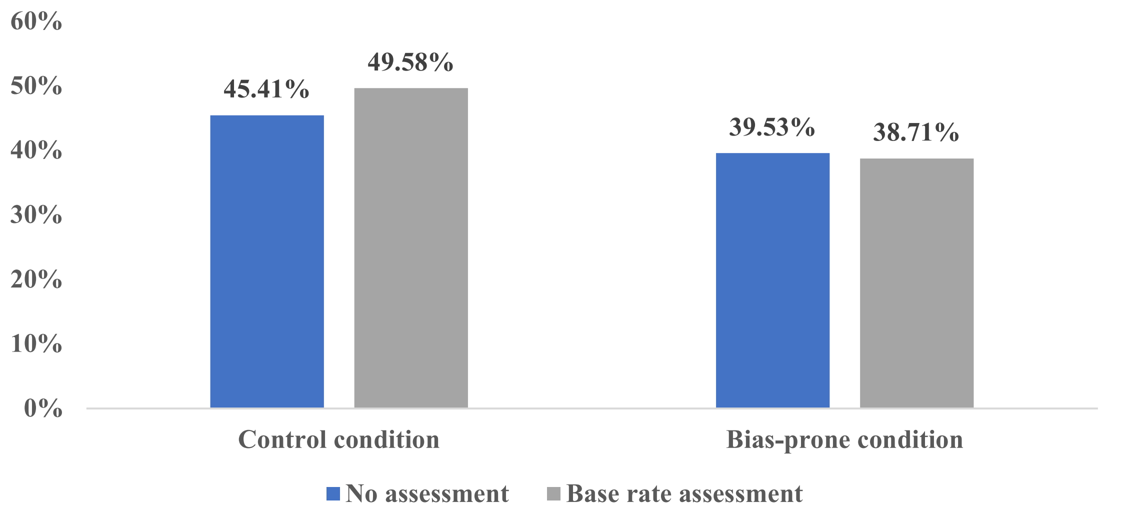
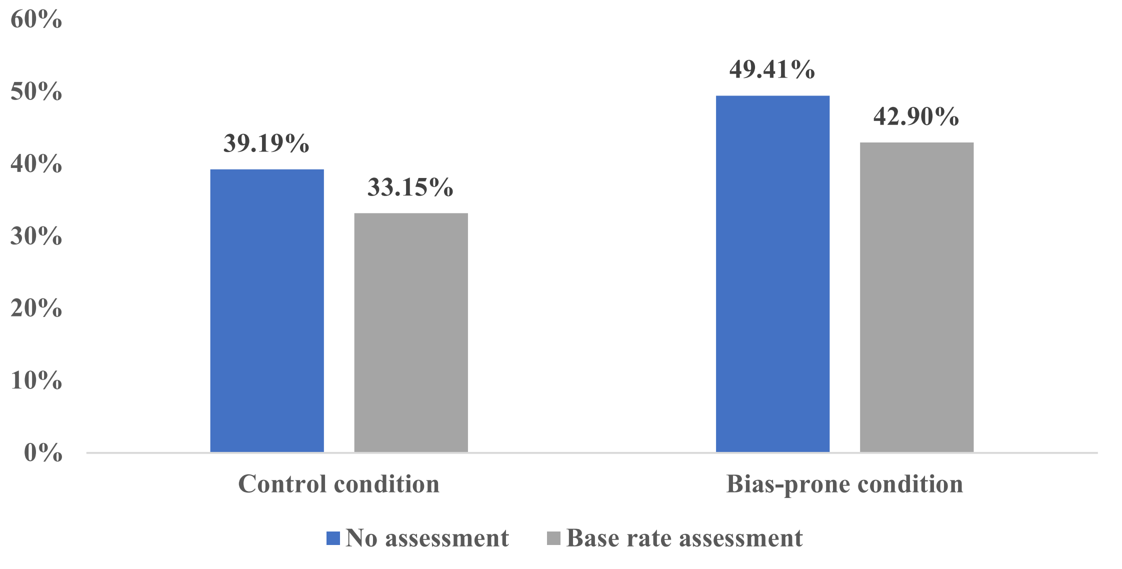


Figure 4 - Percentage of salient disease diagnosis per group



To test H2, we used a logistic regression model having as predictors a binary variable indicating if the participant was assigned to the base rate assessment or not, a binary variable representing the bias-prone condition and their interaction term, in different models for both dependent variables (choice of salient disease and choice of most-likely disease).

As reported in tables 3 and 4, we found no support for an interaction between the bias-prone condition and the base rate assessment as a predictor of choosing the most-likely diagnosis ($\beta = -0.42$, SE = 0.28, $p = 0.13$, Table 3, model 1F), neither predicting the choice of the salient disease ($\beta = 0.07$, SE = 0.25, $p = 0.78$, Table 4, model 2F). Notwithstanding, we found a direct effect of the base rate assessment on choice of the salient disease ($\beta = -0.26$, SE = 0.12, $p < 0.10$, Table 4, model 2E), indicating that it had an effect on the decision, but this effect was not different across the bias-prone and control conditions. Therefore, we found no support for H2.

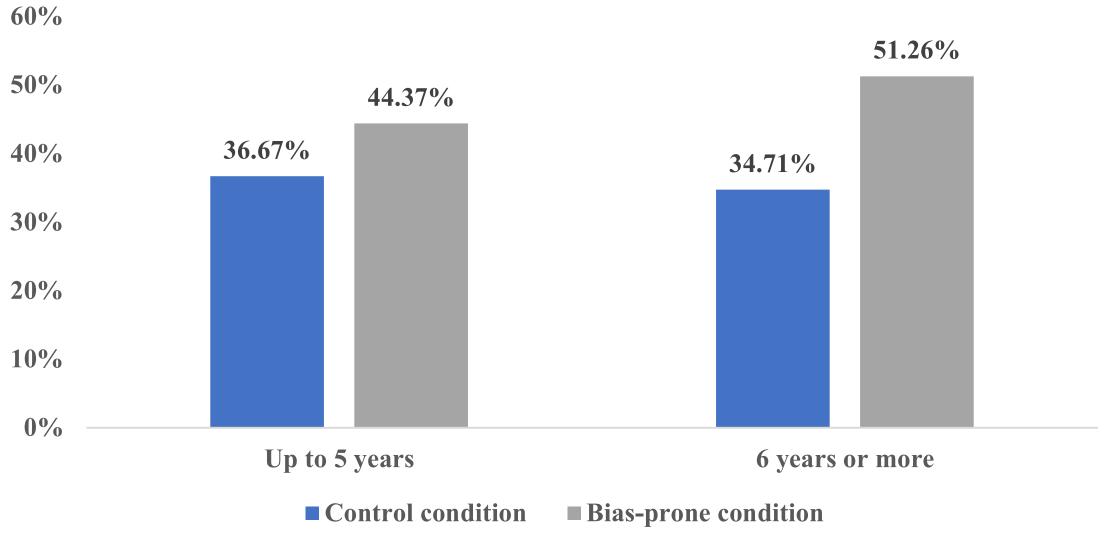
We also found an influence of availability bias in judging the frequency of the salient disease, physicians in the bias-prone condition who took the base rate assessment showed a higher average evaluation of frequency of the salient disease (M = 3.91, SE = 0.44) compared to physicians in control condition who took the base rate assessment (M = 3.66, SE = 0.44, $t(667) = 3.89$, $p < .001$). This result is unexpected, as our exogenous manipulation should not be affected by any characteristics of the participants. It suggests a contamination of our base-rate manipulation by the bias-prone condition, a caveat of our study that may be one reason for the lack of support to H2.

4.4. Exploratory analysis: Clinical experience

To test the role of clinical experience as a moderator of the effect of the bias-prone condition in choosing the salient disease, we estimated logistic regression models including as predictors a binary variable indicating different levels of clinical experience [1 = experienced (6 years or more); 0 = novice (up to 5 years)], a variable representing the bias-prone condition, and their interaction. We found a significant interaction term ($\beta = 0.60$, SE = 0.30, $p = < .05$). In the models without controls, the effect of the bias-prone condition on choice of the salient disease was significant for both groups, while larger among experienced physicians (6 years or more of clinical

experience) ($\beta = 0.68$, SE = 0.24, $p < 0.01$) in comparison to novice physicians (up to 5 years of clinical experience) ($\beta = 0.40$, SE = 0.11, $p < 0.001$, Fig. 5). Interestingly, with the additional controls, the effect loses significance for the novice physicians ($\beta = 0.16$, SE = 0.14, $p = 0.27$), while it remains significant for the experienced participants ($\beta = 0.76$, SE = 0.26, $p < 0.01$). Detailed regression results are included in Appendix C.

Figure 5 - Percentage of salient disease diagnosis by condition and clinical experience



5. DISCUSSION

We found some evidence that supports the effect of availability bias induced by real clinical experience in diagnostic reasoning. While this effect was not significant for the choice of the correct diagnostic in our vignettes (when considering controls), we found support for the effects of the availability bias elicited by real clinical experience with a disease in the last 7 days in increasing the choice of the salient disease, erroneously, as the most probable diagnosis for the case.

This interesting result makes us wonder if availability bias effect takes place only when there is high uncertainty about the right diagnosis. In other words, as availability bias does not significantly promote inaccuracy but increases the choice for the salient disease, we can postulate that those who missed the right diagnosis would miss it anyway, and, due to the high uncertainty about the answer, they were more prone to follow the heuristic.

One could also hypothesize that recent experiences with a disease would increase the knowledge about how to differentiate it from other diseases, a mechanism that could also be playing a role in our experiment, dampening the impact of availability bias. This countervailing force could in part explain why H1 was only partially supported. It could also help explain the heterogeneity in effects across clinical cases (see Appendix C). However, overall, as this mechanism plays against our hypothesis, we found consistent results to believe that salience prevails and explains our findings.

We further hypothesized that performing a base rate assessment of the diseases prior to the diagnosis task would actively mitigate the availability bias (H2) – that is, an exercise of deliberate reasoning (System 2), that would counteract the use of the availability heuristic (System 1). Although the intervention did seem to have a direct effect on increasing diagnostic accuracy and decreasing the choice for the salient disease, this effect was not different between physicians subject to bias and physicians not subject to bias, as the interaction with availability bias was not significant. Therefore, the base rate assessment did improve diagnostic reasoning through turning more evident the base rate as a criteria for diagnosing, but it did not work, as intended, in mitigating the mechanism of availability bias as operationalized in our study. Actually, we found that

bias itself influenced the judgement of frequency, with physicians subject to bias evaluating the salient disease as more frequent in clinical practice than physicians not subject to bias.

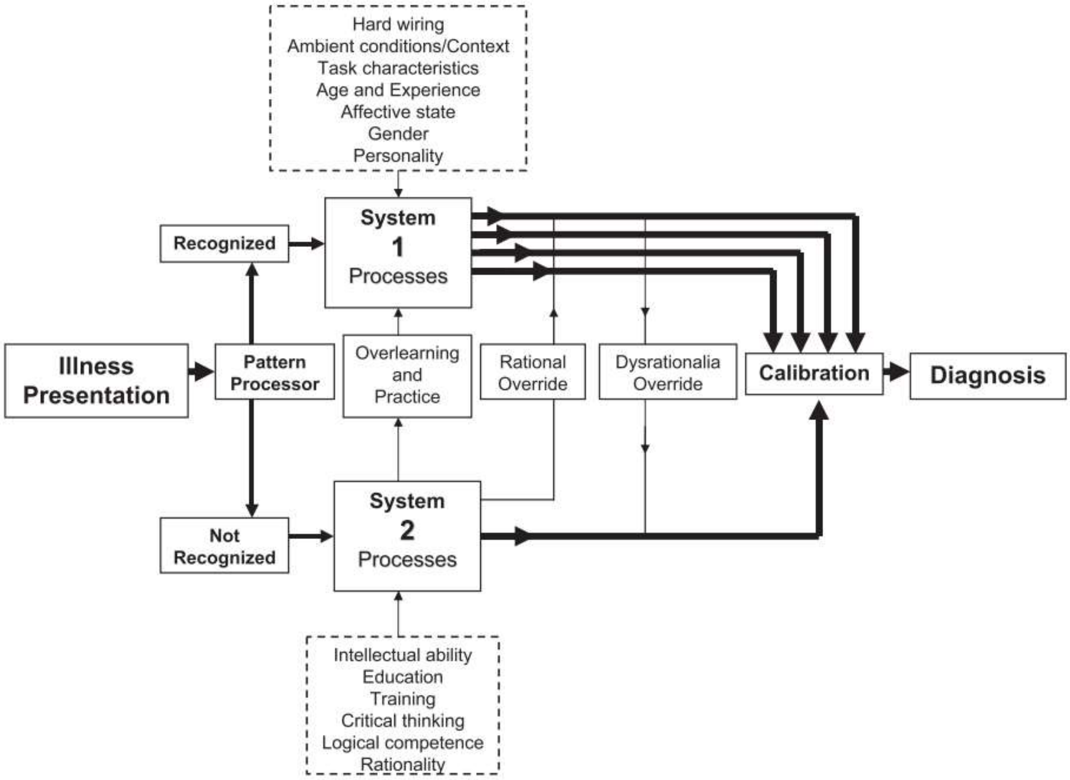
Another interesting finding of our study was that for the first time, in our best knowledge, we found a significant effect of experience in increasing the effect of availability bias comparing novice and experienced physicians. In our sample, experienced physicians were 71% more prone to be biased than novice physicians (up to 5 years of clinical experience). Our results corroborate with several theoretical works attesting that the reliance on non-analytical reasoning tends to increase with experience (Schmidt et al., 1990), as well as it expands the findings of Mamede et al. (2010), who found propensity to availability bias in second-year residents but no propensity in first-year residents, by considering a wider range of experience (novice = up to 5 years vs. experienced = 6 years or more).

Also of notice, as shown in Tables 3 and 4, physicians specialized in the same specialty of the clinical case they evaluated have shown higher diagnostic accuracy and less susceptibility to availability bias. Although this finding seems to contradict our previous claim that reliance on non-analytical reasoning tends to increase with experience, we believe this finding corroborates with other concepts in diagnostic reasoning, particularly overlearning and practice (see Figure 6).

For specialists, besides theoretical knowledge, repetition of similar clinical cases and pattern recognition elicits development of their own clinical rules (a way of System 2 consistently develops System 1 automatic shortcuts) (Croskerry, 2009), which are capable of overriding other System 1 heuristics.

Thus, we shed light in two distinct paths of expertise, in one path we have specialists developing their own clinical rules through overlearning and being less susceptible to availability bias in their own specialties, and, in the other, experienced physicians with no specialization in the area, who does not possess the overlearning of the specialist, although they enjoy the benefits and harms of overconfidence, and, consequently, increased susceptibility to availability bias. Future research should examine if our finding holds for other heuristics and confirm the existence of these distinct paths of expertise.

Figure 6 – Model for diagnostic reasoning



Reproduced from Croskerry, 2009

Although we could not assess the effect of time on availability bias, as our selection process and design did not capture the exact time that the physician attended the patient, we do believe we contribute to literature by showing that, different than previous lab experiments, our bias inducing event happened in a time frame of up to 7 days before the assessment.

6. CONCLUSION

This study presented important advancements in the understanding of availability bias in diagnostic reasoning. We believe that we reached our initial goals in identifying and measuring availability bias effect induced by real experiences in clinical practice. Despite our failure in fully validating our initial hypothesis, more important than this result, by itself, was the understanding of what must come next in our research. Indeed, we found insightful results that will guide further steps of research in availability bias and diagnostic reasoning.

6.1.Contributions

Our greatest advancement with this study was identifying and measuring availability bias induced by real clinical cases, expanding to the field results already established in lab experiments. Despite all noise uncontrolled by our design, that relies in physicians' personal experiences and self-assessment of patient's attendance, we were able to validate the presence of availability bias and its effect in increasing the erroneous choice for the salient disease in a clinical case.

We contribute to the literature on the availability heuristic in diagnostic reasoning in several ways. First, as emphasized above, we bring the concept to real clinical practice. Second, by choosing relatively common diseases as the salient ones, we show that salience is not exclusively a matter of rarity, and therefore common diseases are also able to increase salience, in a way that also promotes greater propensity to biased decisions. Third, although our base rate assessment was unsuccessful in eliminating the availability bias effect, we found that physicians subjected to bias rated the salient disease as more frequent than physicians in control condition, a result consistent with the mechanism of availability bias in increasing the perception of greater likelihood. Fourth, we found a significant availability bias effect in a sample selected by its recent attendance to the salient disease in a 7-day time frame, a much larger interval than the ones used in previous research (in which availability bias was measured immediately after its eliciting event). Finally, we contribute to literature in expertise and decision

making by showing that experienced physicians (with 6 or more years of clinical practice) are more prone to bias than novice physicians, a result predicted by literature review but still lacking empirical findings.

6.2.Limitations and Further Steps

We acknowledge that our study is a small step towards field experiments in clinical reasoning. But what makes our research more interesting, also brings some limitations. In our attempt of relying in physicians real recent experiences as a source of bias, we assigned our groups by participants' self-assessment, what might impose endogeneity into our models, that, despite our efforts of control, might be influencing the results. Further research can better deal with this endogeneity, although we consider it something hard to deal with in designs like ours that gives up the control of lab experiments.

The caveats of our base-rate interventions, as discussed before, also limited our results as it increased noise in the models without aggregating much information to it. In further steps, we should consider remove this intervention to check if new results unshadow.

It is also not possible to affirm if and in which moment during the 7 days of our time frame the physician attended the patient with the salient disease, what limited any of our attempts of assessing for how long availability bias effect holds. Future research should address this limitation as it is an important gap in literature.

As we only asked participants for recent attendance of patients with our disease of interest, we did not have the information about attending the other alternative diseases of the clinical cases. Therefore, we cannot affirm that our physicians were not biased by any of the other diseases. However, we consider that this possibility would play against our results. Even so, this information should be captured in future research.

At last, as mentioned before, in analyzing our data we faced great heterogeneity in our chosen clinical cases (models are presented in Appendix C for appraisal). Although overall results pointed to the expected direction, we acknowledge that there is much more to be discovered in how recent experiences with similar clinical cases influences

further decisions when facing a new one. We actually envisage this heterogeneity and the characteristics of cases that drive it as an opportunity of advancing knowledge in the field and understanding better the different nuances of the effects of attending a patient with a given disease over medical decision making in subsequent patients.

7. REFERENCES

Ariely, D. (2008). *Predictably irrational*. HarperCollins.

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a Cause of Diagnostic Error in Medicine. *American Journal of Medicine*, *121*(5 SUPPL.). https://doi.org/10.1016/j.amjmed.2008.01.001

Blumenthal-Barby, J. S., & Krieger, H. (2015). Cognitive biases and heuristics in medical decision making: A critical review using a systematic search strategy. *Medical Decision Making*, *35*(4), 539–557. https://doi.org/10.1177/0272989X14547740

Bond, W. F., Deitrick, L. M., Arnold, D. C., Kostenbader, M., Barr, G. C., Kimmel, S. R., & Worrilow, C. C. (2004). Using simulation to instruct emergency medicine residents in cognitive forcing strategies. *Academic medicine: journal of the Association of American Medical Colleges, 79*(5), 438–446. https://doi.org/10.1097/00001888-200405000-00014

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of Specific Similarity in a Medical Diagnostic Task. *Journal of Experimental Psychology: General*, *120*(3), 278–287. https://doi.org/10.1037/0096-3445.120.3.278

Croskerry, P. (2009). A universal model of diagnostic reasoning. *Academic Medicine : Journal of the Association of American Medical Colleges*, *84*(8), 1022–1028. https://doi.org/10.1097/acm.0b013e3181ace703

Croskerry, P. (2003). The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*, *78*(8), 775–780. https://doi.org/10.1097/00001888-200308000-00003

Dimara, E., Dragicevic, P., & Bezerianos, A. (2016). *Accounting for Availability Biases in Information Visualization*. 9–11. http://arxiv.org/abs/1610.02857

Elstein, A. S. (2000). Clinical problem solving and decision psychology: comment on “the epistemology of clinical reasoning”. *Academic Medicine : Journal of the Association of American Medical Colleges*, *75*(10 Suppl), 1999–2001. https://doi.org/10.1097/00001888-200010001-00042

Evans J. S. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual review of psychology*, 59, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Graber, M., Gordon, R., & Franklin, N. (2002). Reducing diagnostic errors in medicine: what's the goal?. *Academic medicine: journal of the Association of American Medical Colleges*, 77(10), 981–992. https://doi.org/10.1097/00001888-200210000-00009

Graber M. L. (2013). The incidence of diagnostic error in medicine. *BMJ quality & safety*, 22 Suppl 2(Suppl 2), ii21–ii27. https://doi.org/10.1136/bmjqs-2012-001615

Graber, M. L., Franklin, N., & Gordon, R. (2005). Diagnostic error in internal medicine. *Archives of internal medicine*, 165(13), 1493–1499. https://doi.org/10.1001/archinte.165.13.1493

Institute of Medicine (US) Committee on Quality of Health Care in America, Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (Eds.). (2000). *To Err is Human: Building a Safer Health System*. National Academies Press (US).

Lambe, K. A., O'Reilly, G., Kelly, B. D., & Curristan, S. (2016). Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. *BMJ quality & safety*, 25(10), 808–820. https://doi.org/10.1136/bmjqs-2015-004417

Mamede, S., Mamede, S., De Carvalho-Filho, M. A., De Carvalho-Filho, M. A., De Faria, R. M. D., De Faria, R. M. D., Franci, D., Nunes, M. D. P. T., Ribeiro, L. M. C., Biegelmeyer, J., Zwaan, L., Schmidt, H. G., & Schmidt, H. G. (2020). "Immunising" physicians against availability bias in diagnostic reasoning: A randomised controlled experiment. *BMJ Quality and Safety*, 29(7), 550–559. https://doi.org/10.1136/bmjqs-2019-010079

Mamede, S., van Gog, T., van den Berge, K., Rikers, R. M., van Saase, J. L., van Guldener, C., & Schmidt, H. G. (2010). Effect of availability bias and reflective reasoning on diagnostic accuracy among internal medicine residents. *JAMA*, 304(11), 1198–1203. https://doi.org/10.1001/jama.2010.1276

Mamede, S., Schmidt, H. G., Rikers, R. M., Custers, E. J., Splinter, T. A., & van Saase, J. L. (2010). Conscious thought beats deliberation without attention in diagnostic

decision-making: at least when you are an expert. *Psychological research*, 74(6), 586–592. https://doi.org/10.1007/s00426-010-0281-8

Mamede, S., Goeijenbier, M., Schuit, S., de Carvalho Filho, M. A., Staal, J., Zwaan, L., & Schmidt, H. G. (2021). Specific Disease Knowledge as Predictor of Susceptibility to Availability Bias in Diagnostic Reasoning: a Randomized Controlled Experiment. *Journal of general internal medicine*, 36(3), 640–646. https://doi.org/10.1007/s11606-020-06182-6

Newman-Toker, D. E., & Pronovost, P. J. (2009). Diagnostic errors--the next frontier for patient safety. *JAMA*, 301(10), 1060–1062. https://doi.org/10.1001/jama.2009.249

Norman, G. R., Monteiro, S. D., Sherbino, J., Ilgen, J. S., Schmidt, H. G., & Mamede, S. (2017). The Causes of Errors in Clinical Reasoning: Cognitive Biases, Knowledge Deficits, and Dual Process Thinking. *Academic medicine: journal of the Association of American Medical Colleges*, 92(1), 23–30. https://doi.org/10.1097/ACM.0000000000001421

Ogdie, A. R., Reilly, J. B., Pang, W. G., Keddem, S., Barg, F. K., Von Feldt, J. M., & Myers, J. S. (2012). Seen through their eyes: residents' reflections on the cognitive and contextual components of diagnostic errors in medicine. *Academic medicine: journal of the Association of American Medical Colleges*, 87(10), 1361–1367. https://doi.org/10.1097/ACM.0b013e31826742c9

Reilly, J. B., Ogdie, A. R., Von Feldt, J. M., & Myers, J. S. (2013). Teaching about how doctors think: a longitudinal curriculum in cognitive bias and diagnostic error for residents. *BMJ quality & safety*, 22(12), 1044–1050. https://doi.org/10.1136/bmjqs-2013-001987

Saposnik, G., Redelmeier, D., Ruff, C. C., & Tobler, P. N. (2016). Cognitive biases associated with medical decisions: a systematic review. *BMC Medical Informatics and Decision Making*, 16(1), 1–14. https://doi.org/10.1186/s12911-016-0377-1

Schmidt, H. G., Mamede, S., Van Den Berge, K., Van Gog, T., Van Saase, J. L. C. M., & Rikers, R. M. J. P. (2014). Exposure to media information about a disease can cause doctors to misdiagnose similar-looking clinical cases. *Academic Medicine*, *89*(2), 285–291. https://doi.org/10.1097/ACM.0000000000000107

Schmidt, H. G., Norman, G. R., & Boshuizen, H. P. (1990). A cognitive perspective on medical expertise: theory and implication. *Academic medicine: journal of the Association of American Medical Colleges*, 65(10), 611–621. https://doi.org/10.1097/00001888-199010000-00001

Schwarz, N., & Vaughn, L. A. (2012). The Availability Heuristic Revisited: Ease of Recall and Content of Recall as Distinct Sources of Information. *Heuristics and Biases*, 103–119. https://doi.org/10.1017/cbo9780511808098.007

Sherbino, J., Kulasegaram, K., Howey, E., & Norman, G. (2014). Ineffectiveness of cognitive forcing strategies to reduce biases in diagnostic reasoning: a controlled trial. *CJEM*, 16(1), 34–40. https://doi.org/10.2310/8000.2013.130860

Singh, H., Giardina, T. D., Meyer, A. N., Forjuoh, S. N., Reis, M. D., & Thomas, E. J. (2013). Types and origins of diagnostic errors in primary care settings. *JAMA internal medicine*, 173(6), 418–425. https://doi.org/10.1001/jamainternmed.2013.2777

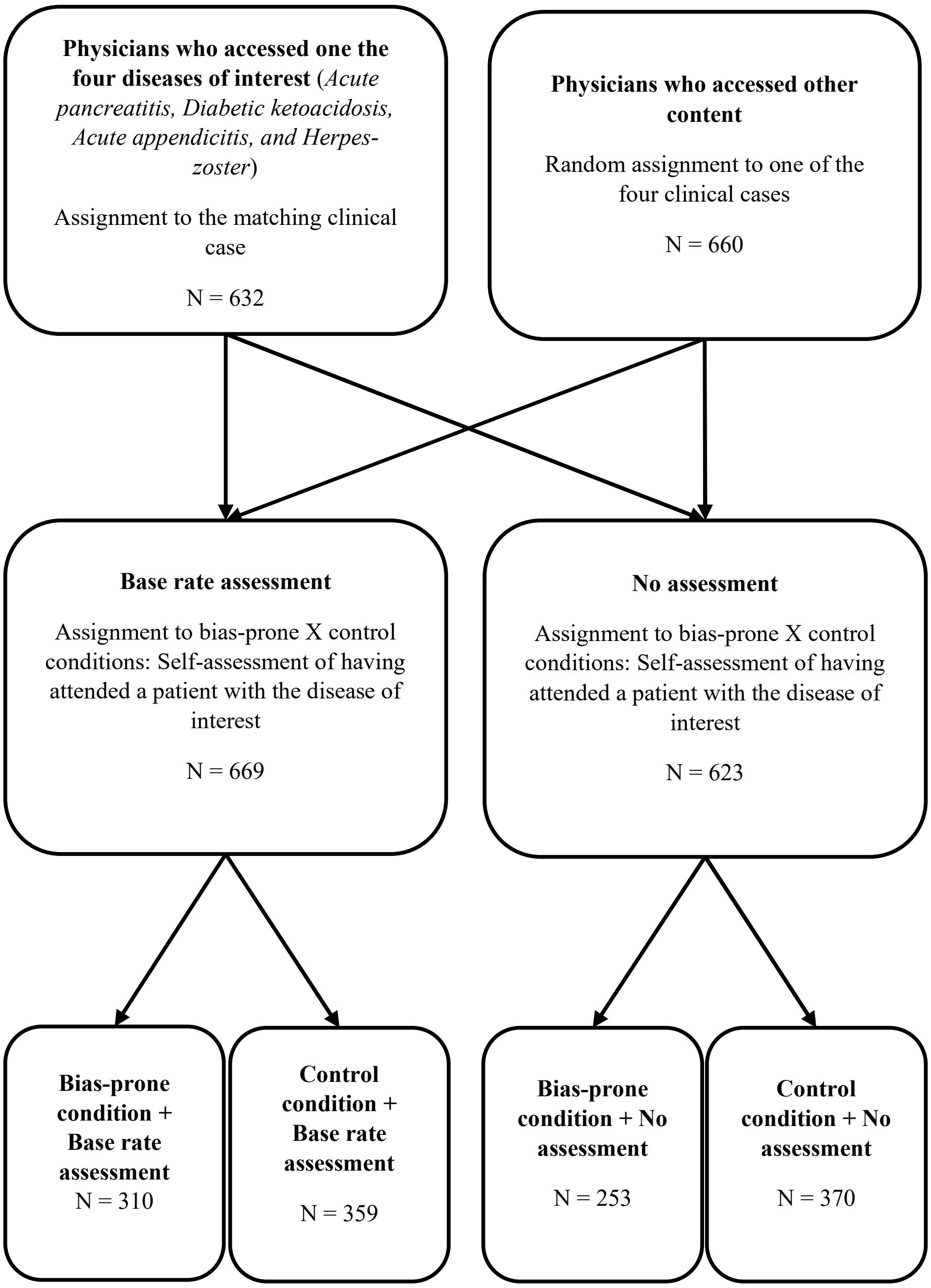
Thaler, R. H. & Sunstein, C.R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science* (New York, N.Y.), 185(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Weingart, S. N., Wilson, R. M., Gibberd, R. W., & Harrison, B. (2000). Epidemiology of medical error. *BMJ* (Clinical research ed.), 320(7237), 774–777. https://doi.org/10.1136/bmj.320.7237.774

APPENDIX A – Selection and assignment process to study conditions



APPENDIX B – Clinical Cases Vignettes

A. Original clinical cases (In Portuguese):

Caso clínico 1 – Especialidades do caso: Gastroenterologia e Cirurgia Geral

Homem de 55 anos se apresenta na unidade de emergência com queixa de dor abdominal nos últimos 3 dias. Ele localiza a dor no andar superior do abdome. Observa que piora depois de comer, causando também náuseas e vômitos. Reclama de febre não aferida e calafrios nas últimas 24 horas.

Apresenta história patológica pregressa de hipertensão e hipercolesterolemia. Nenhuma cirurgia anterior. Não fuma nem usa drogas. Consome bebidas alcoólicas nos fins de semana.

Seus sinais vitais são: PA: 155 x 95 mmHg; FC: 112 bpm; FR: 18 irpm; T: 38,0 C; SpO2 98% em ar ambiente.

Exame físico revela sobrepeso. Exame torácico e cardiovascular são normais, exceto por taquicardia leve. Exame abdominal apresenta sensibilidade à palpação na região epigástrica e hipocôndrio direito, sem sensibilidade de rebote. Peristalse mantida.

<u>Qual o diagnóstico mais provável para este caso?</u>

- Colecistite aguda (Diagnóstico mais provável)

- Pancreatite aguda (Doença saliente)

- Doença ulcerosa péptica

- Trombose de veia porta

Caso clínico 2 – Especialidades do caso: Medicina de Emergência, Geriatria, Medicina Intensiva, e Neurologia

Homem de 74 anos foi trazido à emergência por familiares devido à queda do estado geral e sonolência progressiva nas últimas 72 horas, recusando alimentos e ingestão

hídrica. Paciente com histórico de AVC isquêmico há 5 anos, evoluiu com sequelas motoras que limitavam sua deambulação sem auxílio. Segundo cuidadora, há 1 semana já se encontrava mais prostrado e restrito à cama. Apresenta histórico de hipertensão e diabetes tipo 2.

Sinais vitais: PA: 95 x 57 mmHg, FC: 117 bpm, FR: 27 irpm, Tax = 37,0 C

Ao exame: Rebaixamento do nível de consciência. Desidratação de mucosas e pulso fino. Exame do tórax revela roncos difusos bilateralmente. Exame cardiovascular e abdominal sem alterações significativas. Glicemia capilar à admissão: 267 mg/dL.

<u>Qual o diagnóstico mais provável para este caso?</u>

• Pneumonia bacteriana (Diagnóstico mais provável)

• Cetoacidose diabética (Doença saliente)

• Tuberculose pulmonar

• AVC isquêmico

Caso clínico 3 – Especialidades do caso: Dermatologia

Mulher de 57 anos, obesa, queixa-se de surgimento de lesões cutâneas eritematosas abaixo da mama direita há 6 dias, com muita sensibilidade local, há 24 horas alega o surgimento de vesículas. Trabalha como vendedora ambulante e relata que ao longo do dia prurido e incômodo pioram, principalmente em dias de calor, com intensa sudorese. Nega febre ou outros sintomas. Paciente em uso regular de metformina para controle do diabetes. Nega outras doenças.

Sinais vitais: PA: 134 x 91 mmHg, FC: 87 bpm, FR: 22 irpm, Tax = 36,5 C

Exame físico: Lesões vesiculares em região inframamária direita com intenso rubor local.

<u>Qual o diagnóstico mais provável para este caso?</u>

• Impetigo bolhoso (Diagnóstico mais provável)

• Herpes-zóster (Doença saliente)

• Dermatite de contato

• Foliculite bacteriana

Caso clínico 4 – Especialidade do caso: Cirurgia geral

Homem de 26 anos, sem comorbidades, inicia quadro de dor abdominal em fossa ilíaca direita com irradiação para virilha de forte intensidade há 3 horas, associado a náuseas e vômitos. Nega febre ou outros sintomas.

Sinais vitais: PA: 100 x 67 mmHg, FC: 117 bpm, FR: 23 irpm, Tax = 37,0oC

Exame físico: Dor à palpação profunda em fossa ilíaca direita. Ausência de descompressão dolorosa. Peristalse diminuída. Sem outras alterações.

<u>Qual o diagnóstico mais provável para este caso?</u>

• Litíase urinária (Diagnóstico mais provável)

• Apendicite aguda (Doença saliente)

• Diverticulite aguda

• Apendagite epiploica

B. English translations:

Case 1 – Specialties of the case: Gastroenterology and General Surgery

A 55-year-old man comes to the emergency unit with complaints of abdominal pain in the last 3 days. It locates the pain in the upper abdomen. He notes that he gets worse after eating, also causing nausea and vomiting. He complains about unverified fever and chills in the last 24 hours.

Presents a history of hypertension and hypercholesterolemia. No previous surgery. He does not smoke or use drugs. He consumes alcoholic beverages on the weekends.

His vital signs are: BP: 155 x 95 mmHg; HR: 112 bpm; RR: 18 rpm; T: 38.0 Celsius; 98% SpO_2 in room air.

Physical examination reveals overweight. Thoracic and cardiovascular examination are normal, except for mild tachycardia. Abdominal examination shows sensitivity to palpation in the epigastric region and right hypochondrium, without rebound sensitivity. Peristalsis maintained.

<u>What is the most likely diagnosis for this case?</u>

- Acute cholecystitis (Most-likely diagnosis)
- Acute pancreatitis (Salient disease)
- Peptic ulcer disease
- Portal vein thrombosis

Case 2 – Specialties of the case: Emergency medicine, Geriatrics, Critical care, and Neurology

A 74-year-old man was brought to the emergency by family members due to a drop in general condition and progressive drowsiness in the last 72 hours, refusing food and water intake. The patient has a history of ischemic stroke 5 years ago, having evolved

with motor sequelae that limited his ambulation without assistance. According to caregiver, 1 week ago he was more prostrate and restricted to bed. He has a history of hypertension and type 2 diabetes.

Vital signs: BP: 95 x 57 mmHg, HR: 117 bpm, RR: 27 rpm, T: 37.0 Celsius

On examination: Low level of consciousness. Dehydration of mucous membranes and fine pulse. Examination of the chest reveals diffuse snores bilaterally. Cardiovascular and abdominal examination without significant changes. Capillary blood glucose at admission: 267 mg/dL.

<u>What is the most likely diagnosis for this case?</u>

• Bacterial Pneumonia (Most-likely diagnosis)

• Diabetic Ketoacidosis (Salient disease)

• Pulmonary Tuberculosis

• Ischemic Stroke

Case 3 – Specialty of the case: Dermatology

A 57-year-old obese woman complains about the appearance of erythematous skin lesions below the right breast for 6 days, with great local sensitivity, and in the last 24 hours she claims the appearance of vesicles. She works as a street vendor and reports that throughout the day, itching and discomfort worsens, especially on hot days, with intense sweating. She denies fever or other symptoms. Patient on regular use of metformin to control diabetes. Denies other diseases.

Vital signs: BP: 134 x 91 mmHg, HR: 87 bpm, RR: 22 rpm, T: 36.5 Celsius

Physical examination: Vesicular lesions in the right inframammary region with intense local flushing.

<u>What is the most likely diagnosis for this case?</u>

• Bullous impetigo (Most-likely diagnosis)

• Herpes zoster (Salient disease)

• Contact dermatitis

• Bacterial folliculitis

Case 4 – Specialty of the case: General Surgery

A 26-year-old man, without comorbidities, started abdominal pain in the right iliac fossa with irradiation to the groin of strong intensity for 3 hours, associated with nausea and vomiting. Denies fever or other symptoms.

Vital signs: PA: 100 x 67 mmHg, HR: 117 bpm, FR: 23 irpm, T: 37.0 Celsius

Physical examination: Pain on deep palpation in the right iliac fossa. Absence of painful decompression. Peristalsis decreased. No other changes.

<u>What is the most likely diagnosis for this case?</u>

• Urinary Lithiasis (Most-likely diagnosis)

• Acute Appendicitis (Salient disease)

• Acute Diverticulitis

• Epiploic Appendagitis

APPENDIX C – Supplementary models

Table 5 – Logistic regression models including clinical experience

	DV: Salient diagnosis			
	Model 3A	**Model 3B**	**Model 3C**	**Model 3D**
Bias-prone condition (vs. Control)	0.40***	0.29*	0.32*	0.16
	(0.11)	(0.12)	(0.13)	(0.14)
Experienced physicians (vs. Novice)	0.07	0.15	-0.09	-0.11
	(0.17)	(0.15)	(0.18)	(0.20)
Bias-prone*Experienced physicians			0.36	0.60*
			(0.28)	(0.30)
Specialties match		-0.67†		-0.68†
		(0.41)		(0.41)
Clinical cases				
Case 2		-0.32		-0.31
		(0.20)		(0.20)
Case 3		1.47***		1.50***
		(0.17)		(0.17)
Case 4		0.34		0.36†
		(0.21)		(0.21)
N	1292	1292	1292	1292
Pseudo R-squared	0.01	0.11	0.01	0.11

Standard errors between parentheses

**** $p < .001$, ** $p < .01$, * $p < .05$, † $p < 0.10$*

Table 6 – Logistic regression models for individual clinical cases

	DV: Salient diagnosis							
	Case 1		Case 2		Case 3		Case 4	
	Model 1.1	**Model 1.2**	**Model 2.1**	**Model 2.2**	**Model 3.1**	**Model 3.2**	**Model 4.1**	**Model 4.2**
Bias-prone condition (vs. Control)	0.36	0.42	-0.25	-0.24	0.52**	0.50***	0.12	0.12
	(0.29)	(0.29)	(0.28)	(0.28)	(0.19)	(0.19)	(0.30)	(0.30)
Specialties match		-1.57		-0.87		1.12		-1.64
		(1.06)		(1.06)		(1.09)		(1.07)
N	250	250	336	336	508	508	198	198
Pseudo R-squared	0.01	0.02	0.00	0.01	0.01	0.01	0.00	0.01

Standard errors between parentheses

**** p < .001, ** p < .01, * p < .05, † p < 0.10*