Overconfident health workers provide lower quality healthcare

While a growing body of evidence suggests that healthcare workers in low and middle-income countries often provide poor quality of care, the reasons behind such low performance remain unclear. The literature on medical decision-making suggests that cognitive biases, or failures related to the way healthcare providers think, explain many diagnostic errors. This study investigates whether one cognitive bias, overconfidence, defined as the tendency to overestimate one's performance relative to others, is associated with the low quality of care provided in Senegal. We link survey data on the overconfidence of health workers to objective measures of the quality of care they provide to standardised patients – enumerators who pose as real patients and record details of the consultation. We find that about a third of providers are overconfident – meaning that they overestimate their own abilities relative to their peers. We then show that overconfident providers are 26% less likely to manage patients correctly and exert less effort in clinical practice. These results suggest that the low levels of quality of care observed in some settings could be partly explained by the cognitive biases of providers, such as overconfidence. Policies that encourage adequate supervision and feedback to healthcare workers might reduce such failures in clinical decision-making.

Keywords: cognitive bias; overconfidence; quality of care; medical decision making; overplacement; standardised patients; Senegal.

Classification codes: 2340 (PsycINFO Classification code); D01, I14 (JEL Classification code).

1. INTRODUCTION

Poor quality of care as measured by diagnostic errors or incorrect treatments is worryingly widespread in many settings. Evidence suggests that in the US, doctors incorrectly diagnose between 10% and 15% of their patients (Graber, 2013). Meanwhile, several recent studies show that this problem is even more widespread in low and middle-income countries (LMICs). For instance, in India and China, respectively only 12% and 36% of children suffering from diarrhoea received oral rehydration therapy (Das et al., 2012; Sylvia et al., 2015). Similarly, roughly half of providers in India, Kenya and China failed to recognise the typical signs of pulmonary tuberculosis in their patients and did not order the recommended diagnostic test (Daniels et al., 2017; Das et al., 2015; Sylvia et al., 2017)

The determinants of the low quality of care observed in LMICs are still not well understood. Many studies have examined plausible contributing factors such as providers' working conditions, knowledge, motivation, or inadequate remuneration, but little conclusive evidence has emerged from this body of work (for reviews, see Eijkenaar et al. 2013; Rowe et al. 2005). Meanwhile, the literature on medical decision-making in high-income settings suggests that many diagnostic errors may be due to cognitive biases, or systematic thinking errors that affect providers when they make clinical decisions (Croskerry, 2002; Norman & Eva, 2010). Making decisions about patient care can be a complex process that requires considerable cognitive effort (Elstein & Schwarz, 2002). To make a correct diagnosis and choose the appropriate management for the patient, providers should first listen to the patient's account of their primary symptoms. Then they have to establish all the plausible illnesses from which the patient's symptoms might result and gather further data through systematic questioning and physical examination. Finally, by synthesising and analysing the information obtained, providers determine which condition is the most consistent with the patient's symptoms and decide on the best course of action. Because clinical decision making is often a long and non-linear process, it is prone to error and is especially vulnerable to cognitive biases (Dawson & Arkes, 1987; Elstein, 1999). Previous research has identified various cognitive biases that can lead to errors in providers' judgement (for a review, see Blumenthal-Barby & Krieger 2015). For example, when gathering patient data, providers might only seek information that confirms an initial idea, rather than gather information about other contradictory causes – an issue called confirmation bias (Nickerson, 1998). Similarly, in analysing patient data, availability bias can lead providers to base their judgment on experiences which are more salient – such as recent cases (Mamede et al., 2010).

This study examines the cognitive bias of overconfidence, a concept defined in three different ways in the psychology literature (Moore & Healy, 2008). When overconfidence refers to individuals overestimating the *accuracy* of their judgments, it is called overprecision.[1] When it refers to the idea that people overestimate their *absolute* performance, abilities, or chances of success, it is called overestimation.[2] Finally, when overconfidence refers to respondents overestimating their *relative performance*, i.e. compared to others, it is called overplacement.[3] In this study, we consider healthcare workers' overplacement overconfidence – in reference to their ability to correctly manage hypothetical medical cases. Given the inherent uncertainty in establishing a diagnosis, it is more meaningful to ask providers' to evaluate their performance compared to their peers', rather than in absolute terms. Peer comparisons or benchmarking are also likely to be familiar to providers, given that these are often used in health and ranking is widespread in medical education.

We hypothesize that overconfidence makes providers more likely to mismanage patients in clinical practice, because overestimating one's medical acumen leads to two undesirable behaviours. First, overconfident providers may gather information less thoroughly, as they overestimate the likelihood that their intuition is correct, and therefore underestimate and possibly neglect other plausible causes of illness (McSherry, 1997; Voytovich, Rippey, & Suffredini, 1985). Second, overconfident providers are unlikely to feel the need for assistance or input, in the form of clinical guidelines or advice from peers (Baumann, Deber, & Thompson, 1991; Berner & Graber, 2008).

In this study, we measure the overconfidence of healthcare workers in Senegal and assess the quality of care they provide to patients in clinical practice. We find that about a third of providers interviewed can be characterised as overconfident. In line with our hypothesis, we find that overconfident healthcare workers provider lower quality healthcare, as they are 26% less likely to manage patients correctly. We find that this may be due to the lower effort exerted during consultations, as overconfident providers perform 18% fewer relevant actions (questions and physical examinations) than providers who do not overestimate the quality of their medical judgment.

---

[1] For example, people may overestimate how likely they are to have the correct answers to general knowledge questions (Soll & Klayman, 2004) or how accurate the outcomes of their investment decisions are (Lambert, Bessière, & N'Goala, 2012; Sonsino & Regev, 2013).

[2] For example, when people overestimate how well they will do in exams (De Paola, Gioia, & Scoppa, 2014; Duttle, 2016; Feld, Sauermann, & de Grip, 2017) or how quickly they can complete tasks (Weick & Guinote, 2010).

[3] For example, people often rate their own performance as being above the average in a variety of tasks such as driving or teaching, as well as general knowledge tests (Alicke & Govorun, 2005). People even sometimes think that they know more than experts, as found in a recent study in the USA where a third of respondents thought that their knowledge about the causes of autism exceeded that of medical doctors or scientists (Motta et al., 2018).

This study contributes to the existing literature in three ways. First, in contrast to the current literature on overconfidence in healthcare, which primarily focuses on determining whether or not providers are overconfident (for example, Berner & Graber, 2008; Friedman et al., 2005; Yang et al., 2012a), our study goes one step further and links measures of overconfidence to rich data on provider behaviour in clinical practice. This allows us to test the association between overconfidence and the accuracy of patient management, as well as the level of effort exerted by providers.

Second, this study adds to the literature on overconfidence in healthcare (Croskerry & Norman, 2008; Fortinsky, et al., 2009; Motta, Callaghan, & Sylvester, 2018) by generating evidence from a low-income setting. Existing studies have exclusively focused on measuring overconfidence in patients and healthcare workers in high-income countries – mainly the US and the UK. Given the widespread levels of diagnostic errors in LMICs and the limited understanding of its determinants, this study is a worthwhile addition to the existing literature on the potential effects of cognitive biases in general and overconfidence in particular.

Lastly, this paper contributes to the literature on the determinants of quality of care in LMICs. As mentioned before, there is little consensus on the reasons explaining the high levels of misdiagnosis and inappropriate treatments recently uncovered by audit studies (Das, Woskie, Rajbhandari, Abbasi, & Jha, 2018). The results of this study suggest that psychological factors such as cognitive biases might play a role.

2. DATA AND METHODS

2.1 Study setting and design

Senegal is a lower middle-income country in West Africa, where the quality of primary care is generally poor. Observations of consultations indicate that healthcare workers fail to follow basic clinical protocols, completing less than half of what is recommended in paediatric consultations and 60% in ante-natal consultations (Kruk, Chukwuma, Mbaruku, & Leslie, 2017).

We collected data in a random, representative sample of 197 public health facilities located in Ziguinchor, Sédhiou, Tambacounda, Kédougou – four of the most disadvantaged regions in the country (DHS, 2015). In these areas, public health facilities are the main point of access to modern primary care for the population, as private providers are rare. As is often the case in lower income settings, these facilities are staffed by nurses and midwives (trained between two and three years)

as well as other less or unskilled providers (e.g. auxiliary nurses, nursing trainees or other staff with paramedical or no medical training).

Data collection was organised in two phases. In the first phase (April-July 2016), enumerators visited the sampled facilities to collect information about providers and facilities. 472 health workers present during the facility visits consented to the interviewed and answered a survey that included a clinical knowledge test, as well as a self-assessment of their performance in this test. We also asked providers if they consented to be visited by 'mystery' standardised patients (SPs) in the following six months. SPs are enumerators trained to act like normal patients and consistently present a specific medical case to providers who do not know they are 'fake' patients. Researchers in LMICs are increasingly using SPs to obtain objective measures of the quality of care because of the lack of alternative data sources (Kwan et al., 2019), and evidence suggesting that SPs are the "gold standard" for measuring process quality of care (Luck & Peabody, 2002).

In the second phase of data collection (November-December 2016), teams of five SPs sought to visit all study facilities. As consultations are on a walk-in basis, not all SPs were seen by providers interviewed in phase 1. Hence, of all the providers who received an SP, we could only match N=228 to those interviewed in phase 1. In total, 572 consultations are available to explore the association between overconfidence and quality of care in clinical practice.

Ethical approval for the study was obtained from the research ethics committees of the Senegalese Ministry of Health and the London School of Hygiene and Tropical Medicine.

2.2 Measuring overconfidence

In line with previous studies, we measure provider overconfidence by comparing providers' actual performance and self-assessed performance in a clinical knowledge test (Friedman et al. 2001; Yang et al. 2012b). In our survey, the knowledge test consisted of a series of clinical vignettes in which providers had to diagnose and treat hypothetical patients in a role-playing exercise. An enumerator playing the hypothetical patient started the 'consultation' by describing their main symptom (for example "*I have been coughing for two weeks*"). Providers could then obtain more information by interrogating the 'patient', who would answer following a script. Providers could also indicate which physical examinations or laboratory tests they would perform, and the enumerator would tell them what they would find. The exercise concluded with providers indicating how they would treat the patient. During the role-play, another enumerator recorded all actions taken by the provider: history questions asked, physical examinations and test results requested, advice given and treatments recommended.

Depending on the type of patients seen by providers in the past six months (adults, children and/or family planning)[4], providers assessed up to five vignette cases, in a random order (see Table 1). The cases were based on previous studies using vignettes (Das et al., 2015; Mohanan et al., 2015), and adapted in collaboration with health professionals in Senegal and in the UK. Appendix A.1 provides further details on the clinical cases used.

After completing all clinical vignettes, providers had to indicate for each case if they thought that, compared to their peers, their "level of clinical competence" in the exercise was "below average", "average" or "above average" (see Appendix A2 for details). These responses are used to construct a categorical variable s_{ic} corresponding to provider i's self-assessed performance in case c.

Table 1 – Overview of clinical cases used in the vignettes

Clinical case	Description	Completed by providers who perform	Preferred case management
Dysentery	Child with bloody diarrhoea	Paediatric consultations	ORS & Zinc & Metronidazole
Family planning	Woman asking for a contraceptive to space out births, history of migraines with aura	Family planning consultations	Any contraceptive, except hormonal
Asthma	Respiratory difficulty when coming in contact with dust	Adult consultations	Salbutamol, corticoids or antihistamines
Stable angina	Pressing chest pain (radiating towards neck) following physical activity	Adult consultations	Referral or ECG
Tuberculosis	Persistent cough, blood in sputum, weight loss	Adult consultations	Screening for TB

We use provider i's actual performance in a vignette presenting case c to compute two measures: a_{ic}^e and a_{ic}^m. We construct the first measure a_{ic}^e by using the proportion of relevant actions taken by providers in case c (history questions, physical examinations), and assigning each provider i to one of three groups ("below average", "average" or "above average"), depending on whether their performance places them in the bottom, middle or upper tercile of the distribution.

We compute a_{ic}^m based on whether provider i's management of case c (i.e. treatment and/or tests recommended) is considered incorrect, partially correct or correct, according to national treatment

---

guidelines. Because the distribution of responses is often skewed, to determine if provider i's management of case c is "below average", "average" or "above average" compared to their peers, we compare it to the modal category[5]. For example, in the dysentery and tuberculosis cases, the modal answer for the patient management was "partially correctly". As a result, if respondent i's management of the case was "partially correct", they are categorised as "average"; if their management was incorrect they are categorised as "below average"; and if their management was correct, they are assigned to the "above average" category. Appendix A.3 provides further details about the way each case was coded.

Next, we compute two indicator variables of overconfidence for each vignette: a provider is deemed overconfident in case c if they judged their performance in the vignette task was better than it actually was. Formally, OC_{ic}^k is an indicator variable of individual i's overconfidence in case c and domain k with $k=\{e;m\}$, defined as:

$$OC_{ic}^k = 1 \text{ if } s_{ic} > a_{ic}^k \text{ and } OC_{ic}^k = 0 \text{ if } s_{ic} \leq a_{ic}^k$$

We then aggregate each measure of overconfidence across all cases completed by a provider and construct OC_i^k, an indicator variable that captures provider i's overconfidence in domain k, if OC_{ic}^k =1 in more than half of the cases done. In other words, OC_i^e =1 if provider i's relative performance, in terms of effort during the mock consultation, was lower than they thought and OC_i^m =1 if their relative performance in accuracy of case management was lower than they thought.

Finally, we define a third measure of general overconfidence for provider i as:

$$OC_i = 1 \text{ if } OC_i^e = 1 \text{ and } OC_i^m = 1$$

2.3 Measuring performance in clinical practice

We use data from the SP visits to evaluate how providers perform in clinical practice. The cases portrayed by the five SPs were the same as those used in the vignettes (see Table 1). As indicated before, five SPs visited each sampled facility where any provider could receive them, including some who had not taken part in the baseline survey. As a result, we only have overconfidence data from the baseline survey for N=228 providers, who saw on average 2.5 SPs. SPs were debriefed using a standardised checklist soon after leaving the health facility. The checklist included information on the duration of the consultation, actions done by the provider (questions asked,

---

[5] As detailed in Appendix A.1.6, the quality of case management in vignettes was not symmetrically distributed, so it was not always possible to determine three categories of performance.

physical examination performed), as well as recommendations made and treatments prescribed (see Appendix A.1 for the checklists for each case).

Following previous studies (Daniels et al., 2017; Das, Holla, Mohpal, & Muralidharan, 2016), we use this information to construct two outcomes: whether providers manage the patient correctly and the level of effort providers exert in the consultation. The quality of case management is a binary variable equal to 1 if the patient was managed correctly (i.e. as indicated by national guidelines) and zero otherwise (see Table 1 for definitions of correct case management). Effort is measured by the proportion of history questions asked and physical exams performed during the consultation.

2.4 Measuring provider knowledge

To measure providers' clinical skills i.e. their capacity to correctly manage patients in clinical practice, we construct two measures based on their responses in clinical vignettes. The first measure captures providers' ability to gather relevant information from patients. Using data on the type of questions asked and examinations 'performed' in the vignettes, we use item-response theory (IRT) to create a competence index which is a standardised score capturing their latent knowledge (Das & Hammer, 2005). The second measure reflects provider knowledge of correct treatments and is equal to the proportion of vignettes correctly managed (as in Mohanan et al., 2015).

2.5 Sample characteristics

Table 2 shows some descriptive statistics for sampled healthcare providers and facilities.

Table 2 – Sample characteristics

Variable	Mean	SD	Mean	SD	p-val
Provider characteristics:	*Baseline sample (n=472)*		*Seen by SPs (n=228)*		
Male	0.50	0.50	0.48	0.50	0.41
Is a skilled provider (nurse, doctor, midwife)	0.58	0.49	0.55	0.50	0.23
Doctor	0.05	0.22	0.02	0.13	0.00
Nurse	0.25	0.43	0.26	0.44	0.52
Midwife	0.28	0.45	0.27	0.44	0.71
Other	0.42	0.49	0.45	0.50	0.23
Work experience (years)	9.04	7.73	9.45	8.20	0.27
Is demotivated (intends to quit current job)	0.30	0.46	0.30	0.46	0.91
Was born in the local area	0.24	0.43	0.21	0.41	0.08
Has a permanent contract	0.47	0.50	0.46	0.50	0.75
Attended training on child health (IMCI)	0.42	0.49	0.47	0.50	0.02
Attended training on tuberculosis	0.45	0.50	0.51	0.50	0.01
Attended training on family planning	0.64	0.48	0.74	0.44	0.00
Attended training on non-communicable diseases	0.18	0.38	0.19	0.40	0.41
Clinical knowledge (measured in vignettes):					
Competence index	0.00	1.00	0.02	0.89	0.68
Proportion of vignette cases correctly managed	0.50	0.32	0.51	0.29	0.51
Performance in clinical practice (measured in SPs):					
Proportion of SPs correctly managed	-		0.36	0.48	
Facility characteristics (n=197)	**Mean**	**SD**			
Facility type (1=health post)	0.93	0.26			
Structural quality (% of essential drugs and equipment)	0.78	0.09			
Guidelines (% of treatment guidelines available)	0.52	0.27			
Size of target population	7443.61	6263.57			
Competition (number of facilities in 5km radius)	1.44	2.59			
Distance to closest higher-level facility (km)	37.85	36.90			

Note: p-values refer to t-tests for means and chi-squared tests for proportions.

3. RESULTS

3.1 Are providers overconfident?

In this section, we present descriptive evidence of the prevalence of overconfidence in the sample of 472 providers who completed the vignettes at baseline. We find that health workers generally judge their own performance to be higher than their peers'. Across the 1,726 cases for which they were asked how well they had done compared to their peers, providers thought their performance

was "above average" 39% of the time, "average" 51% of the time and "below average" only 10% of the time.

After comparing their actual performance to their self-assessment and aggregating across vignette cases, we find that 37% of providers can be defined as overconfident in terms of effort undertaken in the consultation (OC_i^e =1). This proportion is slightly lower (32%) if we consider overconfidence in terms of the accuracy of patient management (OC_i^m =1). These two measures are significantly associated (Chi-squared, p=0.000). Overall, 16% of healthcare workers can be defined as generally overconfident (OC_i =1).

3.2 Who is overconfident?

This section explores the correlates of overconfidence. Results in Table 3 show that measures of overconfidence are not consistently associated with provider and facility characteristics. There is some evidence that male providers are 10 percentage points (pp) less likely to be overconfident than women in terms of accuracy of management (column 3). However, once we control for competence in that domain, this association is no longer significant (column 4). We find no gender difference when considering overconfidence in terms of effort undertaken in vignettes.

We find that skilled providers are 20pp less likely to be overconfident in terms of the effort. Tables A1 and A2 in the Appendix includes a more detailed analysis disaggregating by cadres. We find some evidence that doctors are less likely to be overconfident about their effort than other types of providers – although we should be cautious about these results given the small number of doctors in the sample.

A more consistent finding is that overconfidence is negatively associated with provider ability, as providers with higher competence are less likely to be overconfident. A one standard-deviation increase in the competence in information gathering index is associated with a 36pp decrease in the probability of being overconfident (based on measures of effort). Similarly, a 10pp increase in the proportion of vignette cases managed correctly is associated with a 44pp decrease in the probability of being overconfident (defined in terms of correct treatment).

Table 3 – Correlates of provider overconfidence (logistic regressions)

	Overconfidence based on effort ($OC_i^e = 1$)		Overconfident based on management accuracy ($OC_i^m = 1$)	
	(1)	(2)	(3)	(4)
Male	-0.076	0.004	-0.097**	0.011
	(0.050)	(0.036)	(0.046)	(0.046)
Skilled provider	-0.197***	-0.040	-0.069	-0.071
	(0.052)	(0.034)	(0.051)	(0.048)
Work experience	0.000	-0.003*	0.002	0.003
	(0.003)	(0.002)	(0.003)	(0.003)
Demotivated provider	-0.036	-0.042	0.060	0.057
	(0.044)	(0.031)	(0.048)	(0.047)
Provider born in local area	0.000	-0.095**	-0.045	-0.077
	(0.060)	(0.040)	(0.063)	(0.060)
Permanent contract	0.055	0.130***	0.041	0.028
	(0.047)	(0.033)	(0.048)	(0.046)
Distance to higher-level facility	-0.002**	-0.001**	-0.000	-0.000
	(0.001)	(0.000)	(0.001)	(0.001)
Competence index		-0.366***		
		(0.017)		
Proportion of vignette cases correctly managed				-0.440***
				(0.064)
Pseudo R-squared	0.084	0.512	0.051	0.119
Proportion of correct predictions	0.650	0.873	0.686	0.714
Clusters (facilities)	197	197	197	197
Observations (providers)	472	472	472	472

Notes: All models include region and district fixed effects. All models also control for the size of the target population, facility type, facility participation in a results-based financing scheme, structural quality, availability of guidelines and competition. Marginal effects are shown. Standard errors clustered at the facility level in parentheses (**** p<0.01, ** p<0.05, * p<0.1).

3.3 Do overconfident providers deliver worse quality of care?

Using the sample of 228 providers for whom we have overconfidence and SP data, we now examine whether provider overconfidence is associated with the quality of case management in clinical practice. This analysis is conducted at the consultation level with standard errors clustered at the provider level. We estimate the following logistic regression model:

$$(1)\ Y_{icf} = \beta_0 + \beta_1 Overconfident_i + \beta_2 X_i + \beta_3 Z_f + \delta_c + \delta_d + \delta_r + \varepsilon_{icf}$$

where Y_{icf} takes the value 1 if provider i managed the patient presenting case c in facility f according to clinical guidelines. $Overconfident_i$ is a binary variable that captures whether provider i is overconfident. We used the three different definitions of overconfidence defined in the methods

section. We control for a range of provider (X_i) and facility characteristics (Z_f), and include SP case fixed effects (δ_c), as well as district (δ_d) and region (δ_r) fixed effects.

Results in Table 4 show that overconfident health workers are significantly less likely to correctly manage patients in clinical practice. Providers defined as overconfident in terms of accuracy of management of vignettes (OC_i^m) are 9pp (21%) less likely to manage actual patients correctly (see column 2), whilst providers who are defined as generally overconfident (OC_i) are 12pp (26%) less likely to do so (see column 3). However, we do not find evidence that overconfidence in terms of effort invested in vignettes is predictive of the quality of case management in clinical practice.

Table 4 – Overconfidence and the quality of case management (logistic regressions)

	Performance in clinical practice: Case correctly managed		
	(1)	(2)	(3)
Overconfident based on effort (OC_i^e)	-0.030		
	(0.037)		
Overconfident based on management accuracy (OC_i^m)		-0.087**	
		(0.038)	
Overconfident based on both (OC_i)			-0.117***
			(0.044)
Pseudo R-squared	0.142	0.147	0.148
Proportion of correct predictions	0.713	0.720	0.710
Clusters (health providers)	227	227	227
Observations (consultations)	571	571	571

Notes: All models include SP, region and district fixed effects. All models control for facility characteristics (the size of the target population, facility type, facility participation in a results-based financing scheme, structural quality, availability of guidelines and competition) and provider characteristics (gender, skill, experience, motivation, born in local area, permanent contract and training). One observation is dropped in the analysis because it predicts failure perfectly. Marginal effects are shown. Standard errors clustered at the provider level in parentheses (**** p<0.01, ** p<0.05, * p<0.1).

One concern in evaluating the extent to which overconfidence predicts performance in clinical practice is that this association could be confounded by individuals' ability (i.e. their capacity to perform). As shown in Table 3, we find that ability and overconfidence are significantly associated. There are two main reasons for this. The first has to do with the way we measure overconfidence. Since we define overconfidence as the difference between a measure of self-assessed and actual performance (that have only three categories and therefore relatively little granularity), the higher an individuals' ability, the less likely they are to be overconfident. The second is that overconfidence and ability are conceptually linked, as individuals with lower ability are not only worse performers, but are also often worse at assessing their own performance – a phenomenon

called the Dunning-Kruger effect (Dunning and Kruger, 1999). Given that overconfidence and ability are operationally and conceptually linked, in this section we examine whether the association we observe between overconfidence and performance in clinical practice holds when we control for ability.

Given the link between overconfidence and ability, one might have concerns about collinearity when these variables are included in the same model. We test for the degree of correlation between the measures of overconfidence and clinical skills and find that these are not highly correlated (see Tables A3 and A4 in Appendix). Table 5 re-produces the analysis presented in Table 4 including measures of provider ability. As shown in columns 1-3, performance in clinical practice does not seem to be associated with the clinical competence index. However, performance in clinical practice is significantly associated with ability defined as the proportion of vignettes correctly managed (see columns 4-6), as a 10pp increase in the proportion of cases managed correctly in vignettes is associated with a 22pp increase in the probability of correct case management. Importantly, overconfidence remains a significant predictor of the quality of case management, even after controlling for measures of ability. Furthermore, the magnitude of the association between overconfidence and case management in clinical practice only changes marginally when controlling for ability: for overconfident in terms of management accuracy (OC_i^m) the effect size changes from 8.7pp without controlling for ability to 6.5pp when ability is included; for general overconfidence (OC_i), effect sizes change from 11.7pp to 9.4pp.

Table 5 – Ability, overconfidence and the quality of case management (logistic regressions)

	Performance in clinical practice: % of cases correctly managed					
	(1)	(2)	(3)	(4)	(5)	(6)
Competence index	-0.029 (0.031)	0.002 (0.022)	-0.022 (0.023)			
Proportion of vignette cases correctly managed				0.223*** (0.069)	0.197*** (0.073)	0.199*** (0.071)
Overconfident based on effort (OC_i^e)	-0.064 (0.054)			-0.026 (0.037)		
Overconfident based on management accuracy (OC_i^m)		-0.087** (0.038)			-0.065* (0.039)	
Overconfident based on both (OC_i)			-0.131*** (0.047)			-0.094** (0.044)
Pseudo R-squared	0.152	0.154	0.156	0.143	0.147	0.149
Proportion of correct predictions	0.720	0.708	0.717	0.703	0.720	0.715
Clusters (health providers)	227	227	227	227	227	227
Observations (consultations)	571	571	571	571	571	571

Notes: All models include SP, region and district fixed effects. All models also control for facility characteristics (the size of the target population, facility type, facility participation in a results-based financing scheme, structural quality, availability of guidelines and competition) and provider characteristics (gender, skill, experience, motivation, born in local area, permanent contract and training). One observation is dropped in the analysis because it predicts failure perfectly. Marginal effects are shown. Standard errors clustered at the provider level in parentheses (**** p<0.01, ** p<0.05, * p<0.1).

3.4 Do overconfident providers exert less effort in clinical practice?

We now test if overconfidence predicts the level of effort providers invest in clinical practice – which could be one of the pathways through which overconfidence influences the quality of case management. We estimate the following OLS regression:

$$(2) \; Y_{icf} = \beta_0 + \beta_1 Overconfident_i + \beta_2 X_i + \beta_3 Z_f + \delta_c + \delta_d + \delta_r + \varepsilon_{icf}$$

In this case, Y_{icf} is the level of effort provider i exerts in case c in facility f, defined as the proportion of relevant history questions asked and physical examinations performed. The remaining variables are defined as for (1).

We find that overconfident providers conduct consultations less thoroughly. Providers who are overconfident based on effort invested in vignettes perform 6.2pp fewer relevant actions (16%) in clinical practice (see column 1 in Table 6). As shown in column 3, providers who are defined as generally overconfident perform 6.4pp fewer relevant actions (18%) in consultations with patients. We do not find that overconfidence defined in relation to management accuracy is predictive of the level of effort invested in clinical practice.

Table 6 – Overconfidence and provider effort (OLS regressions)

	Performance in clinical practice: Proportion of questions asked and examinations performed		
	(1)	(2)	(3)
Overconfident based on effort (OC_i^e)	-0.062***		
	(0.014)		
Overconfident based on management accuracy (OC_i^m)		-0.020	
		(0.016)	
Overconfident based on both (OC_i)			-0.064***
			(0.018)
R-squared	0.211	0.185	0.203
Clusters (health providers)	228	228	228
Observations (consultations)	572	572	572

Notes: All models include SP, region and district fixed effects. All models also control for facility characteristics (the size of the target population, facility type, facility participation in a results-based financing scheme, structural quality, availability of guidelines and competition) and provider characteristics (gender, skill, experience, motivation, born in local area, permanent contract and training). Marginal effects are shown. Standard errors clustered at the provider level in parentheses (**** p<0.01, ** p<0.05, * p<0.1).

3.5 Robustness checks

Results are robust to a number of alternative specifications.

First, we consider whether results are sensitive to the way we define overconfidence. We aggregate overconfidence across clinical vignettes and define providers as overconfident if they were coded as overconfident in *more than half* of the vignettes they completed. We reproduce the analysis using three different thresholds: an extreme (high) threshold (overconfident in all vignettes), a higher threshold (overconfident in at least three quarters of vignettes) and a lower threshold (overconfident in half or more of vignettes). As expected, we find that associations between overconfidence and quality of case management as well as effort are larger the higher the threshold used, but generally the results hold. Results for the high, higher and lower threshold are respectively shown in Tables A5, A6 and A7.

Second, the internal validity of the SP method relies on providers believing that SPs are real patients. To verify this, we contacted all providers after the SP visits had occurred and asked them to provide details about any patients they suspected to be 'fake'. Of the 98.9% of providers we were able to reach, they correctly identified 46 patients as 'fake' (8% of all consultations). This detection rate is similar to those found in other studies using SPs in rural areas (Sylvia et al., 2015). We next test if our results hold when we exclude those SPs who were detected by providers during the consultation. Tables A8 and A9 show that results are robust to these exclusions.

Third, to address potential concerns for multicollinearity of covariates, we verify whether results hold when facility or provider-level covariates are excluded from the analysis. Tables A10 and A11 shows that our results on the association between overconfidence and behaviour in clinical practice are robust to alternative specifications.

Next, we test whether results still hold if we cluster at the facility level, rather than at the provider level. Tables A12 and A13 reproduces the main results when clustering at the facility level and shows that our findings remain the same.

Table A14 examines whether the patterns of results observed vary by the type of ailment portrayed by SPs and does not find evidence to suggest that this is the case. Finally, as shown in Table A15 and A16, our main results are robust to using OLS and probit regressions

4. DISCUSSION AND CONCLUSION

A body of theoretical work hypothesises that cognitive biases have a negative effect on the quality of care delivered by healthcare workers. In this study, we test the association between a cognitive bias, overconfidence, and rich data on provider behaviour in clinical practice in Senegal. Our results indicate that overconfident healthcare providers are 26% less likely to manage patients correctly.

16

In line with theory, we propose that one plausible mechanism explaining these results is that overconfident healthcare providers conduct consultations less thoroughly – as they complete 18% fewer relevant history questions and physical examinations.

Our findings on the prevalence of overconfidence amongst healthcare providers are in line with previous studies from high-income settings, which measured overconfidence in doctors. For example, Friedman et al., (2005) found that 25% of medical students, 41% of residents and 36% of medical faculty in the US can be classified as overconfident. Despite differences in study setting and design, these levels of overconfidence are comparable to what was found here.

Overall, the magnitude of the correlation we observe between overconfidence and the quality of case management is non-negligible. Previous studies using SPs in LMICs do not find that plausible determinants, such as providers' work experience, their level of clinical knowledge or workload, predict performance (Das et al., 2016; Mohanan et al., 2015). In addition, when studies detect significant association, effect sizes are often much small in magnitude than what we find in relation to overconfidence. For instance, Das et al. (2016) find that providers in India with a medical qualification are only 9% more likely to correctly manage patients than providers without a medical qualification.

This study is limited in several respects. Firstly, we only provide evidence from a relatively small number of healthcare workers, based in rural Senegal. Hence, our sample is not necessarily representative of the population of healthcare workers in Senegal or in LMICs more generally. It is therefore unclear whether findings would hold in another country, in urban areas or amongst more highly skilled providers such as doctors. Secondly, this study offers only some evidence on the channels though which overconfidence might affect outcomes. We explore whether overconfidence is associated with the amount of effort invested (i.e. the proportion of questions asked and physical examinations done). However, it is unclear whether effort is the main channel through which overconfidence might affect the quality of case management – even though this is *a priori* plausible. To be able to examine potential causal pathways in more detail, data on whether providers consult guidelines or consider alternative diagnoses (although hard to obtain) would have been highly informative. Finally, results could be affected by unobservable or unobserved traits. Specifically, provider overconfidence may be associated with personality traits, such as those captured by the "Big Five" questionnaire (Caprara, Barbaranelli, Borgogni, & Perugini, 1993), which have been shown to be associated with overconfidence in other settings (for example, Schaefer, Williams, Goodie, & Campbell, 2004).

Even though we cannot claim that there is a causal link between overconfidence and quality of care provided, taken together with theoretical work on cognitive biases, our results indicate that such a link is plausible. This leads to the question of what can be done to mitigate the potentially detrimental effects of overconfidence. Previous studies in high-income settings have proposed two main strategies for reducing overconfidence in the health setting (Berner & Graber, 2008). The first approach targets individual providers and seeks to get them to reflect more on their clinical decisions, be less subject to biases, more self-aware and better able to recognise when additional information is needed (Croskerry, 2003a). A second approach advocates for system-wide strategies involving better feedback and supervision. These interventions aim to make potential errors more visible to healthcare workers and highlight any mismatch between perceived and actual performance (Duffy & Holmboe, 2006). Previous studies find that feedback can improve performance (Veloski, Boex, Grasberger, Evans, & Wolfson, 2006), although they also highlight that overconfident providers are likely resistant to feedback (Sieck & Arkes, 2005). Whilst the first approach may not necessarily appear as a priority to policymakers in settings where resources are scarce, the second might be more promising and better aligned with other initiatives that are already being implemented in LMICs. In addition to reducing cognitive biases such as overconfidence, evidence suggests that improved supervision and feedback have positive effects on a range of other outcomes such as provider motivation and job satisfaction (Bailey et al., 2016). Further evidence is needed to determine the effect of these strategies on overconfidence.

5. REFERENCES

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In *The self in social judgment* (pp. 85–106). https://doi.org/10.1111/j.1559-1816.2011.00752.x

Bailey, C., Blake, C., Schriver, M., Cubaka, V. K., Thomas, T., & Martin Hilber, A. (2016). A systematic review of supportive supervision as a strategy to improve primary healthcare services in Sub-Saharan Africa. *International Journal of Gynecology & Obstetrics*, *132*(1), 117–125. https://doi.org/10.1016/j.ijgo.2015.10.004

Baumann, A. O., Deber, R. B., & Thompson, G. G. (1991). Overconfidence among physicians and nurses: the "micro-certainty, macro-uncertainty" phenomenon. *Social Science & Medicine (1982)*, *32*(2), 167–174.

Berner, E., & Graber, M. (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

Caprara, G. V., Barbaranelli, C., Borgogni, L., & Perugini, M. (1993). The "big five questionnaire": A new questionnaire to assess the five factor model. *Personality and Individual Differences*, *15*(3), 281–288. https://doi.org/10.1016/0191-8869(93)90218-R

Croskerry, P. (2002). Achieving quality in clinical decision making: cognitive strategies and detection of bias. *Acad Emerg Med*, *9*(11), 1184–1204. https://doi.org/10.1111/j.1553-2712.2002.tb01574.x

Croskerry, Pat. (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

Croskerry, Pat, & Norman, G. (2008). Overconfidence in Clinical Decision Making. *American Journal of Medicine*, *121*(5 SUPPL.). https://doi.org/10.1016/j.amjmed.2008.02.001

Daniels, B., Dolinger, A., Bedoya, G., Rogo, K., Goicoechea, A., Coarasa, J., … Das, J. (2017). Use of standardised patients to assess quality of healthcare in Nairobi, Kenya: a pilot, cross-sectional study with international comparisons. *BMJ Global Health*, *2*(2), e000333.

https://doi.org/10.1136/bmjgh-2017-000333

Das, J, Holla, A., Das, V., Mohanan, M., Tabak, D., & Chan, B. (2012). In Urban And Rural India, A Standardized Patient Study Showed Low Levels Of Provider Training And Huge Quality Gaps. *Health Affairs*, *31*(12), 2774–2784. https://doi.org/10.1377/hlthaff.2011.1356

Das, J, Holla, A., Mohpal, A., & Muralidharan, K. (2016). Quality and Accountability in Healthcare Delivery: Audit-Study Evidence from Primary Care in India. *American Economic Review*, *106*(12), 3765–3799. https://doi.org/10.3386/w21405

Das, J, Kwan, A., Daniels, B., Satyanarayana, S., Subbaraman, R., Bergkvist, S., … Pai, M. (2015). Use of standardised patients to assess quality of tuberculosis care: A pilot, cross-sectional study. *The Lancet Infectious Diseases*, *15*(11), 1305–1313. https://doi.org/10.1016/S1473-3099(15)00077-8

Das, J, Woskie, L., Rajbhandari, R., Abbasi, K., & Jha, A. (2018). Rethinking assumptions about delivery of healthcare: implications for universal health coverage. *BMJ (Clinical Research Ed.)*, *361*, k1716. https://doi.org/10.1136/bmj.k1716

Das, Jishnu, & Hammer, J. (2005). Which doctor? Combining vignettes and item response to measure clinical competence. *Journal of Development Economics*, *78*(2), 348–383. https://doi.org/10.1016/J.JDEVECO.2004.11.004

Dawson, N. V., & Arkes, H. R. (1987). Systematic errors in medical decision making: - Judgment limitations. *Journal of General Internal Medicine*, *2*(3), 183–187. https://doi.org/10.1007/BF02596149

De Paola, M., Gioia, F., & Scoppa, V. (2014). Overconfidence, omens and gender heterogeneity: Results from a field experiment. *Journal of Economic Psychology*, *45*, 237–252. https://doi.org/10.1016/j.joep.2014.10.005

DHS. (2015). *Senegal 2015 DHS Final Report*. http://dhsprogram.com/what-we-do/survey/survey-display-489.cfm.

Duffy, F. D., & Holmboe, E. S. (2006, September 6). Self-assessment in lifelong learning and improving performance in practice: Physician know thyself. *Journal of the American Medical Association*, *296*(9), 1137–1139. https://doi.org/10.1001/jama.296.9.1137

Duttle, K. (2016). Cognitive skills and confidence: interrelations with overestimation, overplacement and overprecision. *Bulletin of Economic Research*, *68*(S1), 42–55.

https://doi.org/10.1111/boer.12069

Eijkenaar, F., Emmert, M., Scheppach, M., & Schöffski, O. (2013). Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy*, *110*(2–3), 115–130. https://doi.org/10.1016/j.healthpol.2013.01.008

Elstein, A. S. (1999). Heuristics and biases: selected errors in clinical reasoning. *Academic Medicine : Journal of the Association of American Medical Colleges*, *74*(7), 791–794. https://doi.org/10.1097/00001888-199907000-00012

Elstein, A., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *British Medical Journal*, *324*(March), 729–732.

Feld, J., Sauermann, J., & de Grip, A. (2017). Estimating the relationship between skill and overconfidence. *Journal of Behavioral and Experimental Economics* , *68*, 18–24. https://doi.org/10.1016/j.socec.2017.03.002

Fortinsky, R. H., Panzer, V., Wakefield, D., & Into, F. (2009). Alignment between balance confidence and fall risk in later life: Has over-confidence been overlooked? *Health, Risk & Society*, *11*(4), 341–352. https://doi.org/10.1080/13698570903015735

Friedman, C., Gatti, G., Elstein, A., Franz, T., Murphy, G., & Wolf, F. (2001). Are clinicians correct when they believe they are correct? Implications for medical decision support. *Studies in Health Technology and Informatics*. https://doi.org/10.3233/978-1-60750-928-8-454

Friedman, C. P., Gatti, G. G., Franz, T. M., Murphy, G. C., Wolf, F. M., Heckerling, P. S., … Elstein, A. S. (2005). Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. *Journal of General Internal Medicine*, *20*(4), 334–339. https://doi.org/10.1111/j.1525-1497.2005.30145.x

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

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kruk, M. E., Chukwuma, A., Mbaruku, G., & Leslie, H. H. (2017). Variation in quality of primary-care services in Kenya, Malawi, Namibia, Rwanda, Senegal, Uganda and the United Republic of Tanzania. *Bull World Health Organ*, *95*, 408–418. https://doi.org/10.2471/BLT.16.175869

Kwan, A., Daniels, B., Bergkvist, S., Das, V., Pai, M., & Das, J. (2019). The use of standardized patients for health care quality research in low- and middle-income countries. *BMJ Global Health*, *(forthcomi*.

Lambert, J., Bessière, V., & N'Goala, G. (2012). Does expertise influence the impact of overconfidence on judgment, valuation and investment decision? *Journal of Economic Psychology*, *33*(6), 1115–1128. https://doi.org/10.1016/J.JOEP.2012.07.007

Luck, J., & Peabody, J. W. (2002). Using standardised patients to measure physicians' practice: validation study using audio recordings. *British Medical Journal*, *325*(September), 1–5.

Mamede, S., Van Gog, T., Van den Berge, K., Rikers, R., Van Saase, J., Van Guldner, C., & Schmidt, H. (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

McSherry, D. (1997). Avoiding premature closure in sequential diagnosis. *Artificial Intelligence in Medicine*, *10*(3), 269–283.

Mohanan, M., Vera-Hernández, M., Das, V., Giardili, S., Goldhaber-Fiebert, J. D., Rabin, T. L., … Seth, A. (2015). The know-do gap in quality of health care for childhood diarrhea and pneumonia in rural India. *JAMA Pediatrics*, *169*(4), 349–357. https://doi.org/10.1001/jamapediatrics.2014.3445

Moore, D. A. D. A., & Healy, P. J. P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502

Motta, M., Callaghan, T., & Sylvester, S. (2018). Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes. *Social Science and Medicine*, *211*(June), 274–281. https://doi.org/10.1016/j.socscimed.2018.06.032

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Norman, G. R., & Eva, K. W. (2010). Diagnostic error and clinical reasoning. *Medical Education*, *44*(1), 94–100. https://doi.org/10.1111/j.1365-2923.2009.03507.x

Rowe, A., De Savigny, D., Lanata, C., & Victora, C. (2005). How can we achieve and maintain high-quality performance of health workers in low-resource settings? *Lancet*, *366*(9490), 1026–1035.

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the Big Five. *Journal of Research in Personality*, *38*(5), 473–480. https://doi.org/10.1016/j.jrp.2003.09.010

Sieck, W. R., & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making*, *18*(1), 29–53. https://doi.org/10.1002/bdm.486

Soll, J. B., & Klayman, J. (2004). Overconfidence in Interval Estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 299–314. https://doi.org/10.1037/0278-7393.30.2.299

Sonsino, D., & Regev, E. (2013). Informational overconfidence in return prediction – More properties. *Journal of Economic Psychology*, *39*, 72–84. https://doi.org/10.1016/J.JOEP.2013.06.006

Sylvia, S., Shi, Y., Xue, H., Tian, X., Wang, H., Liu, Q., … Rozelle, S. (2015). Survey using incognito standardized patients shows poor quality care in China's rural clinics. *Health Policy and Planning*, *30*(3), 322–333. https://doi.org/10.1093/heapol/czu014

Sylvia, S., Xue, H., Zhou, C., Shi, Y., Yi, H., Zhou, H., … Das, J. (2017). Tuberculosis detection and the challenges of integrated care in rural China: A cross-sectional standardized patient study. *PLoS Medicine*, *14*(10). https://doi.org/10.1371/journal.pmed.1002405

Veloski, J., Boex, J. R., Grasberger, M. J., Evans, A., & Wolfson, D. B. (2006). Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7. *Medical Teacher*, *28*(2), 117–128. https://doi.org/10.1080/01421590600622665

Voytovich, A. E., Rippey, R. M., & Suffredini, A. (1985). Premature conclusions in diagnostic reasoning. *Journal of Medical Education*, *60*(4), 302–307.

Weick, M., & Guinote, A. (2010). How long will it take? Power biases time predictions. *Journal of Experimental Social Psychology*, *46*(4), 595–604. https://doi.org/10.1016/j.jesp.2010.03.005

Yang, H., Thompson, C., & Bland, M. (2012a). Effect of improving the realism of simulated clinical judgement tasks on nurses' overconfidence and underconfidence: Evidence from a comparative confidence calibration analysis. *International Journal of Nursing Studies*, *49*(12), 1505–1511. https://doi.org/10.1016/j.ijnurstu.2012.08.005

Yang, H., Thompson, C., & Bland, M. (2012b). The effect of clinical experience, judgment task difficulty and time pressure on nurses' confidence calibration in a high fidelity clinical simulation. *BMC Medical Informatics and Decision Making*, *12*(1), 113. https://doi.org/10.1186/1472-6947-12-113