

## Differences between self-reported and laboratory measures of diabetes, chronic kidney disease, and hypercholesterolemia

Pedro Cisalpino Pinheiro (<https://orcid.org/0000-0002-6954-1708>)<sup>1</sup>

Marilisa Berti de Azevedo Barros (<https://orcid.org/0000-0003-3974-195X>)<sup>2</sup>

Célia Landmann Szwarcwald (<https://orcid.org/0000-0002-7798-2095>)<sup>3</sup>

Ísis Eloah Machado (<http://orcid.org/0000-0002-4678-2074>)<sup>4</sup>

Deborah Carvalho Malta (<http://orcid.org/0000-0002-8214-5734>)<sup>5</sup>

**Abstract** *This paper aims to compare the self-reported prevalence measured by laboratory tests and the false positive and negative values for diabetes, chronic kidney disease, and hypercholesterolemia. We used information from the interview and laboratory tests of the National Health Survey (2013, 2014-2015). Sensitivity and specificity were calculated by gender, age, schooling, having health insurance, and time since the last medical visit. We used logistic regression to analyze associated factors with false positives and negatives. Sensitivity was higher for diabetes and among older adults and those who had a medical visit more recently. Specificity was high for all diseases, with better performance among younger people, those with high schooling, and a visit more than one year ago. The likelihood of false positives and negatives decreased with schooling and increased with age. Low sensitivity suggests that prevalence might be higher than indicated by self-reported measures.*

**Key words** *Sensitivity and specificity, Self-report, Diabetes Mellitus, Chronic kidney disease, Hypercholesterolemia*

<sup>1</sup> Faculdade de Medicina, Universidade Federal de Minas Gerais (UFMG). Av. Professor Alfredo Balena 190, Santa Efigênia. 30130-100 Belo Horizonte MG Brasil.

[pedrocalpino@gmail.com](mailto:pedrocalpino@gmail.com)

<sup>2</sup> Faculdade de Ciências Médicas, Unicamp. Campinas SP Brasil.

<sup>3</sup> Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fiocruz. Rio de Janeiro RJ Brasil.

<sup>4</sup> Departamento de Medicina de Família, Saúde Mental e Coletiva, Universidade Federal de Ouro Preto. Ouro Preto MG Brasil.

<sup>5</sup> Escola de Enfermagem, UFMG. Belo Horizonte MG Brasil.

## Introduction

The self-reported health measure is often used as an alternative to surveys that, in general, require a more complicated data collection process and entail higher costs<sup>1</sup>. However, Moreira *et al.*<sup>1</sup> emphasize that self-reported prevalence (often found in household surveys) may be biased. Respondents may classify themselves as ill and not carry the disease (false positive) or not report the disease and be diagnosed with the disease (false negative).

Some factors that may explain the lack of understanding about the health situation itself. For example, Johnston *et al.*<sup>2</sup> highlight that individuals do not recognize their condition. They may have provided incorrect information to doctors, or forgotten or misinterpreted medical advice, or even received incorrect information from professionals. The same authors point out that the lack of understanding about the health condition itself can vary by socioeconomic level. Thus, the analysis of self-reported health conditions will probably be influenced by factors such as income and education<sup>3</sup>.

According to Velakkal *et al.*<sup>4</sup>, epidemiological studies on chronic NCDs in developing countries tend to show lower prevalence in groups with lower socioeconomic status, mainly due to the difficulty of accessing health services. In India, the authors compared the self-reported diagnosis with the diagnosis based on standard tests (structured questionnaires for identifying some diseases) for some NCDs and identified essential differences. Self-reported diagnoses were more prevalent in groups with higher income and education. However, the differences were not maintained when the standard test was used, which, in the authors' opinion, could be associated with the difficulty of groups in worse socioeconomic situations in accessing health services, even if they can identify their worst health condition.

In Brazil, considering the population over 18, the prevalence of self-reported diabetes mellitus (DM) was higher than that diagnosed from laboratory tests, according to PNS data (2014-2015)<sup>5</sup>. The difference between the prevalence of DM was more significant for the group of 60 and over, and the self-reported measure was more prevalent. Laboratory diagnoses evidenced a higher prevalence than self-reported in the 35-44 years age group.

Regarding chronic kidney disease (CKD), laboratory estimates (PNS 2014-2015) show that the number of individuals with the disease is approximately four times higher than the self-re-

ported<sup>6</sup>. Concerning cholesterol, the situation is very similar to that observed with CKD. The prevalence measured by laboratory tests is considerably higher than what that pointed out by the self-reported diagnoses<sup>7</sup>.

Diabetes, high cholesterol, and CKD have different symptomatic and clinical manifestations, which implies different probabilities of their identification by individuals<sup>8-10</sup>. The pathological profile influences sociodemographic characteristics, access to health services, and how people are aware of their health<sup>8-10</sup>. However, the joint analysis – for different diseases – of the differences between self-reported and laboratory-measured diagnoses can help us better measure the quality of self-reported health information and the differences between the outcomes. The analysis stratified by different sociodemographic variables can also be of great importance for the planning of public health policies when pointing out possible, more sensitive groups. The National Health Survey (PNS) contains a series of biological markers that make the proposal feasible.

In this context, this study aims to compare the prevalence between the self-reported diagnosis – collected in the first stage of the National Health Survey (PNS), in 2013 – and that measured through laboratory tests in the second stage of the research, in the 2014-2015 period, based on sensitivity and specificity, for diabetes, chronic kidney disease, and high cholesterol. These measures will be analyzed considering sociodemographic characteristics and access to health services. We also intend to analyze how these characteristics influence the probability of false-positive and false-negative diagnoses.

## Methods

This is a cross-sectional study using information from the two stages of the National Health Survey (PNS). The self-reported data were constructed from the answer to the question if any doctor had already diagnosed them with some of the diseases of interest. A specific question for each of the outcomes is included in the PNS chronic disease module collected in 2013, namely, Diabetes (Q030), Cholesterol (Q060), and Chronic Kidney Disease (Q124). These outcomes were compared with laboratory tests of Glycosylated Hemoglobin (GH), Creatinine, Glomerular Filtration Rate (GFR), Total and Fractional Cholesterol (LDL, HDL, and TC), performed on the same individuals between 2014 and 2015.

The base with PNS laboratory exams consists of 8,952 observations, from a subsample of the PNS. The construction of the PNS laboratory base can be better understood in other studies<sup>5,11-13</sup>. The number of tests analyzed here ranged from 7,211 to 8,528 due to losses in the laboratory sample, processing of analyses, insufficient material, and others. The sample considered for each of the outcomes, and each of the different metrics had observations with self-reported and laboratory information. That is, cases with missing values in any of these variables were disregarded.

The laboratory diagnosis of diabetes was based on glycosylated hemoglobin (HgA1) for individuals with HgA1  $\geq$  6.5. This cutoff point was used in a recently published work with the same database<sup>5</sup>. It should be noted that cases where individuals had HgA1  $\leq$  6.5 (which would represent the absence of diabetes) but declared having used medication to lower sugar or insulin in the two weeks before the research (Q03401 or Q03402) were excluded from the analysis on diabetes. Keeping these cases in the base could artificially inflate false-positive cases and specificity. The same exercise could not be performed for the other outcomes due to the availability of information.

Two different metrics were used for Chronic Kidney Disease (CKD): the glomerular filtration rate (GFR) and creatinine (CR). The definition of cutoff points followed the work of Malta et al.<sup>6</sup>, with the diagnosis of CKD attributed to men with CR  $\geq$  1.3 and 1.1 for women. All those with values below 60 mL/min/1.73 m<sup>2</sup> were considered to have CKD based on the GFR. The calculation of the GFR in the database was based on the equation of the MDRD study without adjustment for ethnicity and skin color<sup>14</sup>. Three different metrics and cutoff points were employed to define the diagnosis of altered cholesterol: total cholesterol  $\geq$  200 mg/dL; low-density lipoproteins (LDL)  $\geq$  130 mg/dL; and high-density lipoproteins (HDL)  $<$  40 mg/dL for men and  $<$  50 mg/dL for women. It should be noted that the question used for self-reported identification refers to high (and not altered) cholesterol, which could lead to some confusion concerning HDL. The definition of limits was based on a study with the same database<sup>7</sup> and a Brazilian guideline on dyslipidemia<sup>15</sup>.

The first strategy for analyzing the differences between laboratory-measured and self-reported diagnoses was the assessment of sensitivity and specificity for each of the outcomes of interest,

considering the set of sociodemographic variables – gender, age, schooling, whether or not they have health insurance, and the time elapsed since the last medical appointment. The sensitivity indicates the proportion of individuals who reported having one of the diseases among those that the laboratory test indicated the disease's presence. In turn, specificity is the proportion of individuals who responded that they did not have the disease, and the test confirmed its absence<sup>16</sup>; that is, it indicates the proportion of cases in which the laboratory measurement confirmed the negative self-reported diagnosis. In the analysis, the laboratory base's outcomes were addressed as the gold standard. The sensitivity and specificity of each of the categories of variables of interest for each outcome and metric were calculated separately. Thus, confidence intervals (95% CI) were generated for each subgroup.

Then, logistic regression models were developed for two dependent variables (for each outcome), false-positive and false-negative cases. That is, for diabetes, CKD, and cholesterol, the false-positive variable was assigned a value of "1" in cases where individuals declared the presence of the disease, but the information was not corroborated by the laboratory test, and zero for the other cases. The value of "1" in the false-negative variable was assigned to individuals whose laboratory diagnosis indicated the presence of the disease, not confirmed by the respondents, and a value of zero was attributed in the other cases. The same sociodemographic and access to health services variables described in the sensitivity and specificity analysis were used as explanatory variables. The model was developed for the three outcomes considering each of the metrics used (HgA1, GFR, CR, TC, LDL, and HDL).

The 2013 PNS was approved by the National Research Ethics Committee (CONEP) of the National Health Council, Ministry of Health. The research participants signed an informed consent form (ICF), and subsequently, collections of peripheral blood were performed at any time of the day. All estimates and analyses were made using Data Analysis and Statistical Software (Stata), version 2014.

## Results

### Diabetes

The prevalence of diabetes measured based on glycosylated hemoglobin (HgA1) in the

sample analyzed here was 7.4% (95% CI, 6.7%-8.2%), while the self-reported information points to a prevalence of 5.5% (95% CI, 4.9% -6.3%). The difference between the prevalence levels was statistically significant. Table 1 shows that 59% (95% CI, 53.8%-63.4%) of individuals with the laboratory diagnosis of diabetes declared to have the disease.

Men and women sensitivity to diabetes was practically the same, at 58.9% (95% CI, 50.5%-66.9%) and 59% (95% CI, 52.4%-65.2 %), respectively. Considering the age groups, the sensitivity concerning diabetes was lower in the younger group (less than 50 years), and the difference between this group and the others (50-59 and 60 and over) was statistically significant. Schooling levels did not show significant differences in diabetes-related sensitivity.

Individuals who had visited the doctor over a year earlier, with a laboratory diagnosis of diabetes, reported the presence of the disease (26.8% - 95% CI, 17.0%-39.5%) at levels lower than those groups whose medical visit occurred more recently (both above 60% sensitivity). A health insurance plan did not show a significant difference in sensitivity for laboratory and self-reported diabetes measurements.

Table 2 shows the specificity of self-reported measures and based on laboratory tests for all outcomes. Concerning diabetes, 98.8% (95% CI, 98.4%-99.1%) of individuals with glycosylated hemoglobin below 6.5% reported not having the disease. In general, the specificity of diabetes was relatively high. Among the variables analyzed here, only age showed a significant difference between the subgroups. Specificity for people under 50 years (99.3% 95% CI, 98.9%-99.6%) was higher than among those aged 60 and over (97%).

Tables 3 and 4 show, respectively, for each outcome, the odds ratio estimated by a logistic regression model for two dependent variables, namely, false-negative (FN) and false-positive (FP) cases. Considering diabetes, the likelihood of false-negatives increased significantly with age. Regarding the group aged 50 or less, both individuals aged 50-59 years (OR = 2) and those older than 60 years (OR = 3.18) were more likely to show FNs. Regarding schooling, the model shows that individuals with access to higher education were 51% less likely (OR = 0.49) to record a false negative than those in the illiterate or incomplete elementary group. The other variables (gender, last visit, and health insurance plan) did not show statistical significance in the model estimated to analyze FNs' for diabetes.

Table 4 indicates that being over 60 significantly increased (183%, OR = 2.83) the likelihood of a false positive occurrence for diabetes compared to the group under 50. The other variables were not statistically significant.

### Chronic Kidney Disease

The prevalence of creatinine-based chronic kidney disease was 5% (95% CI, 4.5%-5.6%) and 6.6% (95% CI, 6-7.4%) considering the Glomerular Filtration Rate. In turn, the self-reported prevalence of CKD was 1.4% in both samples used to analyze outcomes. The sensitivity of laboratory and self-reported measurements was relatively low for both GFR and creatinine, with 4.4% (95% CI, 2.7%-7%) and 3.2% (95% CI, 2.1 %-5.1%), respectively. That is, a small portion of individuals with the laboratory diagnosis of CKD adequately reported the presence of the disease. Considering both metrics, the sensitivity analysis of chronic kidney disease (Table 1) did not show significant differences between the analyzed variables' subgroups. However, noteworthy are the low levels of sensitivity observed for all subgroups.

As for specificity, Table 2 shows high values (98.7%, 95% CI, 98.3%-99%) for both metrics (Creatinine and GFR). In conjunction with diabetes, it was the highest specificity observed among the outcomes. As in the sensitivity analysis, no significant differences were identified between the subgroups of sociodemographic variables and access to health services.

Table 3 shows that the likelihood of FN was higher for the group over 60 (OR = 3.63) compared to the group under 50, considering creatinine as a laboratory measure. In the same model, having a health insurance plan increased the likelihood of FN for CKD by 40% (OR = 1.4). The analysis of the occurrence of FN for CKD – when GFR was the metric used – points to a higher likelihood of FN for women (OR = 1.65). Individuals from both the 50-59 years group (OR = 3.35) and the 60 years and over group (OR = 6.95) were more likely to declare that they were not sick, and the statement diverged from the GFR. The same model indicated a lower likelihood of false-negative (OR = 0.73) for those with a medical visit more than a year earlier, compared to the group whose visit took place in less than three months.

Table 4 shows statistically significant differences in the likelihood of false-positive only for the 50-59 years group compared to those under 50, and only in the creatinine-based model. None

**Table 1.** Self-reported sensitivity of Diabetes Mellitus, CKD, and cholesterol (TC, HDL, LDL) compared to laboratory measurements.

	Diabetes	Chronic Kidney Disease (Creatinine)	Chronic Kidney Disease (GFR)	Total Cholesterol	Cholesterol LDL	Cholesterol HDL
Total	59	4.4	3.2	24.2	27.4	16.5
IC 95%	(53.8;63.9)	(2.7;7)	(2.1;5.1)	(22.2;26.3)	(24.6;30.3)	(15.1;17.2)
N.Obs.	7347	8528	7450	7222	7222	7211
Gender						
Women	59	3.7	2.6	27.1	30.3	18.5
95% CI	(52.4;65.2)	(1.9;7.3)	(1.3;5)	(24.6;29.9)	(26.7;34)	(16.6;21.4)
N.Obs.	4469	4980	4340	4428	4427	4422
Men	58.9	5	4.4	19.8	23.3	13.4
95% CI	(50.5;66.9)	(2.6;9.4)	(2.4;8.1)	(16.8;23.3)	(19;28.1)	(11.2;15.9)
N.Obs.	2878	3548	3110	2794	2795	2789
Age						
Less than 50	42.2	2.8	3.3	16.5	18	10.5
95% CI	(31.1;54)	(0.9;7.9)	(1.2;8.8)	(14;19.4)	(14.5;22)	(8.9;13.1)
N.Obs.	4238	5019	4756	4029	4028	4022
50-59	65.9	5	2.4	27.8	33.5	21.3
95% CI	(56.2;74.4)	(2;12.1)	(1;5.8)	(23.6;32.5)	(27.6;39.9)	(17.8;25.2)
N.Obs.	1314	1495	1419	1318	1318	1318
60 and over	64(57;70.5)	5.4	3.7	33	36.4	26.7
95% CI	(57;70.5)	(2.9;9.9)	(2;6.5)	(29.3;37)	(31.4;41.7)	(23.6;26.7)
N.Obs.	1795	2014	1275	1875	1876	1871
Schooling						
Illiterate or Incomplete Elementary	61.2	6.9	4.9	27.3	31.8	20.5
95% CI	(54.7;67.3)	(3.9;11.8)	(2.8;8.3)	(24.4;30.5)	(27.9;36.1)	(18.2;20.6)
N.Obs.	3113	3785	3086	3087	3088	3079
Complete Elementary or Incomplete High School	54.6	1.8	1.3	21	23.2	13.5
95% CI	(45;63.8)	(0.8;4.3)	(0.5;3.2)	(18;24.5)	(19;28)	(11.4;16)
N.Obs.	3079	3499	3217	2982	2981	2979
Higher Education	62.1	1.9	2	23.8	26	14.9
95% CI	(44.6;77)	(0.4;8.3)	(0.6;6.7)	(19.1;29.3)	(19.2;34.1)	(11.7;21.3)
N.Obs.	1155	1244	1147	1153	1153	1153
Last visit						
Less than 3 months	64.4	5.9	3.4	27.8	31.7	19.7
95% CI	(57.8;70.5)	(3.5;10)	(1.9;6)	(25;30.8)	(27.9;35.9)	(17.7;20.9)
N.Obs.	3958	4437	3790	3970	3971	3961
3 months to 1 year	60.8	1.6	3	26.4	30.5	17.1
95% CI	(50.4;70.3)	(0.4;6.2)	(1.1;7.9)	(22.2;31.1)	(24.6;37.1)	(14.2;19.8)
N.Obs.	1770	1982	1755	1733	1732	1733
Over 1 year	26.8	2.1	2.9	10.8	11.9	6.7
95% CI	(17;39.5)	(0.6;7.3)	(1.1;7.7)	(8.3;14)	(8.5;16.4)	(4.9;11.1)
N.Obs.	1619	2109	1905	1519	1519	1517
Health Insurance Plan						
No	56.5	4.1	3.5	22.6	26.4	15.4
95% CI	(50.5;62.2)	(2.3;7.1)	(2.1;6)	(20.3;25)	(23.3;29.8)	(13.8;16.9)
N.Obs.	5480	6527	5732	5335	5335	5327
Yes	65.4	4.9	2.6	27.8	29.8	19.1
95% CI	(55;74.5)	(2.1;10.9)	(1.1;5.9)	(23.9;32.1)	(24.4;35.9)	(16.2;20.8)
N.Obs.	1867	2001	1718	1887	1887	1884

Source: National Health Survey (PNS), 2013; 2014-2015.

**Table 2.** Self-reported specificity of Diabetes Mellitus, CKD, and Cholesterol (TC, HDL, LDL) compared to laboratory measurements.

	Diabetes	Chronic Kidney Disease (Creatinine)	Chronic Kidney Disease (GFR)	Total Cholesterol	Cholesterol LDL	Cholesterol HDL
Total	98.8	98.7	98.7	89.1	87.5	85.7
95% CI	(98.4;99.1)	(98.3;99)	(98.3;99)	(87.9;90.2)	(86.4;88.6)	(84.2;87.1)
N.Obs.	7347	8528	7450	7222	7222	7211
Gender						
Women	98.6	98.5	98.5	87.7	85.6	83.3
95% CI	(98.1;99)	(97.9;98.9)	(97.8;99)	(86;89.2)	(84;87)	(81.1;85.2)
N.Obs.	4469	4980	4340	4428	4427	4422
Men	99	99	99	90.8	89.9	88.1
95% CI	(98.4;99.4)	(98.5;99.3)	(98.5;99.3)	(89;92.3)	(88.2;91.3)	(85.9;90)
N.Obs.	2878	3548	3110	2794	2795	2789
Age						
Less than 50	99.3	99	99	94.2	93	93.2
95% CI	(98.9;99.6)	(98.5;99.3)	(98.5;99.3)	(93;95.3)	(91.8;94)	(91.6;94.4)
N.Obs.	4238	5019	4756	4029	4028	4022
50-59	98.5	98	98.1	84.4	83.5	79.1
95% CI	(97.5;99)	(96.9;98.7)	(97;98.8)	(80.5;87.7)	(80.2;86.3)	(74.6;83)
N.Obs.	1314	1495	1419	1318	1318	1318
60 and over	97.3	98.4	98.3	76.7	75.4	71.6
95% CI	(96.1;98.1)	(97.6;99)	(97.2;99)	(73.4;79.7)	(72.6;78.1)	(67.7;75.1)
N.Obs.	1795	2014	1275	1875	1876	1871
Schooling						
Illiterate or Incomplete Elementary	98	98.8	98.9	85	83.7	80.8
95% CI	(97.3;98.5)	(98.3;99.1)	(98.5;99.2)	(82.8;86.9)	(81.8;85.5)	(78.2;83.2)
N.Obs.	3113	3785	3086	3087	3088	3079
Complete Elementary or Incomplete High School	99.2	98.4	98.4	92.5	90.7	90
95% CI	(98.5;99.5)	(97.7;98.9)	(97.7;99)	(90.8;93.9)	(89.1;92.1)	(87.8;91.8)
N.Obs.	3079	3499	3217	2982	2981	2979
Higher Education	99.3	99.3	99.2	88.8	87.2	85
95% CI	(98.6;99.7)	(98.5;99.7)	(98.2;99.7)	(85.6;91.3)	(84.3;89.5)	(81.2;88.2)
N.Obs.	1155	1244	1147	1153	1153	1153
Last visit						
Less than 3 months	98.4	98.5	98.5	86.3	84.6	82.6
95% CI	(97.9;98.8)	(97.9;98.9)	(97.8;98.9)	(84.5;87.9)	(83;86.1)	(80.4;84.7)
N.Obs.	3958	4437	3790	3970	3971	3961
3 months to 1 year	99.2	98.7	98.8	89.6	87.6	85.2
95% CI	(98.6;99.6)	(97.9;99.2)	(98;99.3)	(87.1;91.7)	(85.2;89.6)	(81.7;88)
N.Obs.	1770	1982	1755	1733	1732	1733
Over 1 year	99.2	99.2	99.2	96.3	95.4	94.7
95% CI	(98.1;99.7)	(98.5;99.6)	(98.4;99.6)	(94.5;97.5)	(93.8;96.6)	(92.7;96.2)
N.Obs.	1619	2109	1905	1519	1519	1517
Health Insurance Plan						
No	98.6	98.8	98.9	90.3	88.9	87.3
95% CI	(98.1;99)	(98.4;99.1)	(98.5;99.2)	(88.9;91.5)	(87.7;90.1)	(85.6;88.8)
N.Obs.	5480	6527	5732	5335	5335	5327
Yes	99.2	98.5	98.4	86.8	84.7	82.9
95% CI	(98.6;99.5)	(97.5;99.1)	(97.3;99.1)	(84.3;88.9)	(82.4;86.7)	(79.9;85.6)
N.Obs.	1867	2001	1718	1887	1887	1884

Source: National Health Survey (PNS), 2013; 2014-2015.

**Table 3.** Estimated Odds ratio for the occurrence of False Negative, by socioeconomic variables, for each outcome analyzed, PNS 2013, 2014-2015.

	Diabetes	Chronic Kidney Disease (Creatinine)	Chronic Kidney Disease (GFR)	Total Cholesterol	Cholesterol LDL	Cholesterol HDL
N° Obs.	7347	8528	7450	7222	7222	7211
Gender						
Men	-	-	-	-	-	-
Women	1.3	0.79	1.65*	1.15	1.11	1.39*
Age						
Less than 50	-	-	-	-	-	-
50-59	2*	1.3	3.35*	1.69*	1.58*	0.8*
60 and over	3.18*	3.63*	6.95*	1.41*	1.37*	0.8*
Schooling						
Illiterate or Incomplete Elementary	-	-	-	-	-	-
Complete Elementary or Complete High School	0.92	0.95	0.83	0.99	1.01	1.02
Higher Education and over	0.49*	0.75	0.83	0.91	0.89	0.82
Last visit						
Less than 3 months	-	-	-	-	-	-
3 months to 1 year	1.15	0.91	0.82	1.11	1.07	1.1
1 year and over	1.18	0.75	0.73*	1.31*	1.41*	1.21*
Health Insurance Plan						
Yes	0.78	1.4*	1.2	0.89	0.88	0.75*
No	-	-	-	-	-	-

Note: \* Statistical significance of 0.95.

Source: National Health Survey (PNS), 2013; 2014-2015.

of the explanatory variables showed statistical significance in the GFR-based analysis.

### Cholesterol

The laboratory base shows cholesterol above the recommended of 33.9% (95% CI, 32.4%-35.3%) and 19.5% (95% CI, 18.3%-20.7%) considering total cholesterol (TC) and LDL, respectively. In turn, the tests indicate that 49.8% (48.3% -51.4%) of the sample had HDL below the recommended. The prevalence of self-reported high cholesterol was 15.4%.

The sensitivity (Table 1) related to cholesterol was considerably lower than that observed in the analysis of diabetes. The sensitivity was higher for LDL, at 27.4% (95% CI, 24.6%-30.3%), followed by total cholesterol (24.2%, 95% CI,

22.2%-26.3%) and HDL (16.5%, 95% CI, 15.1%-17.2%). TC and HDL showed a higher sensitivity for women. In the three cholesterol measurements, the group's sensitivity under 50 was at least half of those aged 60 and over. No statistical significance was observed in any of the three cholesterol metrics among the schooling subgroups.

Table 1 indicates that the time elapsed since the last visit influences the sensitivity for practically all outcomes. The same was observed in analyses based on TC, LDL, and HDL, particularly when comparing the groups that had visited a doctor more than one year earlier and the group that had visited a doctor in the last three months since the interview.

Specificity was lower for the three cholesterol metrics concerning diabetes and CKD, ranging from 85.7% (95% CI, 84.2%-87.1%) and 89.1%

**Table 4.** Estimated odds ratio for the occurrence of False Positive, by socioeconomic variables, for each outcome analyzed, PNS 2013, 2014-2015.

	Diabetes	Chronic Kidney Disease (Creatinine)	Chronic Kidney Disease (GFR)	Total Cholesterol	Cholesterol LDL	Cholesterol HDL
N° Obs.	7347	8528	7450	7222	7222	7211
Gender						
Men	-	-	-	-	-	-
Women	1.27	1.36	1.32	1.14	1.32*	1.05*
Age						
Less than 50	-	-	-	-	-	-
50-59	1.92	1.97*	1.76	1.97*	2.04*	3.17*
60 and over	2.83*	1.47	1.46	2.98*	3.15*	3.93*
Schooling						
Illiterate or Incomplete Elementary	-	-	-	-	-	-
Complete Elementary or Complete High School	0.7	1.46	1.6	0.73*	0.8	0.77
Higher Education and over	0.61	0.58	0.69	1.02	1.02	1.21
Last visit						
Less than 3 months	-	-	-	-	-	-
3 months to 1 year	0.54	0.93	0.84	0.81	0.89	0.94
1 year and over	0.6	0.63	0.64	0.33*	0.36*	0.4*
Health Insurance Plan						
Yes	0.66	1.21	1.36	1.32*	1.37*	1.47*
No	-	-	-	-	-	-

Note: \* Statistical significance of 0.95.

Source: National Health Survey (PNS), 2013; 2014-2015.

(95% CI, 24.6%-30.3%) (Table 2). Specificity measured by HDL and LDL was higher for men. Table 2 shows that the age gradient of specificity shows the opposite sign to that observed in the sensitivity (Table 1). In other words, specificity decreased with age. For all outcomes, the younger population (less than 50) had a higher proportion of non-sick (laboratory-confirmed) individuals who reported not having the disease than the 60 years and over group. For example, in the case of HDL, the absolute difference in specificity between these groups was 21.6%.

In turn, considering schooling, the differences were only statistically significant between illiterates or those with incomplete primary education and those with complete primary or secondary education. Specificity was higher for individuals who went to the doctor over a year earlier. Measured by LDL, specificity was higher for those who did not have health insurance.

The likelihood of false-negative (Table 3) was higher for women when measured by HDL (OR = 1.39). Regarding the younger group (under 50 years), the likelihood of FN was almost always positive with age. The likelihood of false-negative decreased with age only for HDL. Concerning time since the last visit, the likelihood of recording FNs was higher for TC, LDL, and HDL, respectively, with a 31% (OR = 1.31), 41% (OR = 1.41), and 21% (OR = 1.21) higher likelihood (1 year or more) compared to those who had gone to the doctor less than three months earlier. In turn, having a health insurance plan reduced the likelihood of FNs only for cholesterol measured by HDL (OR = 0.75).

Table 4 shows that women were 32% more likely to declare having been diagnosed with high cholesterol and having LDL within the expected (laboratory-confirmed) (FP) (OR = 1.32) than men. The likelihood of FP for women was also



higher, considering HDL (OR = 1.05). The likelihood of someone over 60 years of age reporting FP was greater for TC (OR = 2.98), LDL (OR = 3.15), and HDL (OR = 3.93).

Few outcomes attested that the schooling level affected the likelihood of FP. The likelihood of someone with elementary school or complete high school (total cholesterol) representing a case of PF was 27% less likely than illiterates or with incomplete elementary school. The longer the time elapsed since the last visit, the lower the likelihood of FP, especially for the group in which the last visit to the doctor occurred over a year earlier. Having a health insurance plan pointed to a higher likelihood of FPs.

## Discussion

Considering the metrics used, the results indicate low agreement between self-reported and laboratory-measured diagnoses. Biological markers can inform about health conditions before the individual's perception of symptoms<sup>3</sup>. Crimmins and Vasunilashorn<sup>17</sup> highlight that the availability of biomarkers allows researchers to know the health situation and the effectiveness of services simultaneously. The difference between self-reported and laboratory-measured diagnoses, especially concerning sensitivity, indicates an essential proportion of the population that may be ill and is unaware of this condition.

PNS laboratory data show that chronic kidney disease estimates were up to four times higher than self-reported data<sup>7</sup>. The results showed that sensitivity was higher in diabetes, where about 60% of those who claimed to have the disease were positively tested. Concerning cholesterol, this measure was lower (close to a quarter) and less than 5% for CKD. Therefore, the low sensitivity identified here may reflect underdiagnosed CKD, especially in the country. However, the comparison between the measures of agreement between different diseases should be performed with caution. Self-reported health measurements' accuracy is related to socioeconomic and demographic characteristics, the presence or absence of morbidities, and the pathology's features<sup>8-10</sup>. The perception of symptoms is one of the factors related to disease recognition<sup>8-10</sup>.

Diabetes was more sensitive than other diseases. To a large extent, this result may be related to the manifestation of the disease's typical symptoms, not observed in cholesterol and CKD. Okura et al.<sup>10</sup> note that diabetes is not an aggres-

sive disease at the onset of its manifestation, but requires many interactions with the health system for its control, which increases the likelihood of individuals recognizing the disease. Compared to CKD and high cholesterol, this may be one reason for the increased sensitivity observed. While it remains asymptomatic for a long time, cholesterol is more commonly requested in routine tests than CKD. The clinical manifestation of the condition increases the likelihood of individuals reporting it<sup>9</sup>.

In this sense, Harris and Schoorp<sup>18</sup> highlight that, unlike self-reported measures, biometric markers offer objective health measures and can point out risks in cases where symptoms do not manifest early and consistently. This reinforces the importance of research, such as the PNS, for planning health actions at the national level.

Specificity, or the proportion of individuals who responded that they did not have the disease and the test confirmed it, was high for all measures (diabetes, CKD, and cholesterol), with a lower cholesterol level. Higher specificity was found in younger individuals with a high schooling level and who had visited a doctor over a year earlier. However, these same groups have a lower prevalence of the diseases analyzed<sup>5-7</sup>.

The current study confirms the literature results that point to low sensitivity and high specificity (96%) between measured and self-reported measures<sup>10,19-22</sup>. Analyzing diabetes and hypertension, Ning et al.<sup>16</sup> observed a higher sensitivity for the older population. A cohort study on older adults in Bambuí also identified a higher sensitivity for diabetes among older adults who visited the doctor most recently and among those with some schooling<sup>19,23</sup>. In general, sensitivity was higher among older people.

In most of the outcomes and metrics considered, the group that had visited a doctor less than three months earlier had a higher sensitivity than those with a last visit in a more extended period. Greater access to health services, with a higher likelihood of having the disease diagnosed, and the ability to remember medical guidelines and change attitudes can explain this outcome<sup>19,23</sup>. Other studies also point out that the use of health services is the factor most associated with the validity of self-reported morbidity<sup>17,19,24-26</sup>.

As for education, some studies point to different directions. In China, Wu et al.<sup>20</sup> identified a greater ability to correctly report the disease among those with higher education and a higher number of chronic diseases. In turn, Goldman et al.<sup>22</sup> associate the quality of self-reported infor-

mation with better cognitive function. Another study in China on diabetes showed that the group with primary education had a higher sensitivity than secondary education<sup>16</sup>. In the Dutch context, Molenaar *et al.*<sup>21</sup> observed a higher sensitivity in the intermediate level of schooling than those with higher education. The PNS showed greater sensitivity in the groups with lower education (illiterate or incomplete elementary school), especially when compared to the intermediate level (complete elementary or complete high school). In general, the group that accessed the upper level showed higher sensitivity than the intermediate level but lower than the first group.

However, it should be noted that this indicator is very likely to suffer from the effect of age composition. Lower schooling levels are more common in the older population, the same group that has higher sensitivity. In this sense, the results are influenced by the age structure of the education groups. Considering only the population aged 60 and over with diabetes, to illustrate the argument, sensitivity shows the expected gradient, 74.7% (Higher Education and over), 65.8% (Elementary or High School), and 62.7% (Illiterate or Incomplete Elementary). In turn, in general terms, specificity grows with the schooling level, agreeing with other studies<sup>16,21</sup>. Concerning the composition effect on specificity, we can also understand that the age groups with the highest schooling level are, in general, younger, the same group with the highest specificity<sup>27</sup>.

Education and access to health insurance plans can be considered income proxies. Dallas *et al.*<sup>3</sup> emphasize that, unlike self-reported health measurements, biological markers may reveal an association between disease and income before the illness becomes explicit. Harris and Schorpp<sup>18</sup> emphasize that identifying risks from biological markers allows planners to intervene in the social determinants of health risk before the disease manifests itself or the condition worsens.

The regression models reinforce the impression that age is a crucial factor in analyzing the agreement between measured and self-reported diagnoses. The likelihood of incorrectly declaring the diagnosis, for all outcomes, increased substantially with age for both false-positives (individuals without a laboratory diagnosis, but who self-reported the disease's presence) and for false-negatives (positive laboratory diagnosis, not endorsed by the self-reported information). This result is interesting in light of the higher sensitivity for the older age groups. While having proportionally greater knowledge about their

health condition, older adults are also very likely to misreport their condition, regardless of the outcome analyzed.

This relationship can be identified in both directions in the literature. For example, Johnson *et al.*<sup>2</sup> observe a positive relationship between age and the occurrence of false-negative for chronic hypertension. In turn, Onur and Valmuri<sup>28</sup> observed a negative effect of age on the occurrence of FN for hypertension and pulmonary disease. It should be noted that these authors used models and specifications different from those used here. Compared to the less educated group, in general, the likelihood of FP or FN was not statistically significant. In India, Onur and Valmuri<sup>28</sup> note that higher schooling levels reduce the propensity to record FN.

The declining likelihood of false-positives as the time since the last visit increases is probably related to patients' low ability to understand the diagnoses. Velakkal *et al.*<sup>4</sup> emphasize that one of the factors behind the discrepancy between self-reported measures and laboratory measures is often the inability to understand the diagnosis received. More vulnerable individuals (older, less educated) have lower levels of knowledge about their health.

It should be noted that false-positive cases (and specificity) can be artificially influenced by cases of individuals with the medication-controlled disease. In these cases, the self-reported diagnosis will be positive (correctly), but the laboratory measurement will point to another direction due to disease control. Likewise, some previous tests may have returned a positive result at another time in life and, even if the parameters are no longer altered, individuals may report the presence of the disease. Concerning diabetes, the database allowed us to exclude from the analysis cases in which diabetes could be medication-controlled. However, the same exercise could not be performed for the other outcomes. Another situation that must be considered is the possibility of PNS testing errors and potential errors in other previous tests. In general, health services usually repeat tests to clear doubts.

This work has significant limitations. The period between the two data collections, on average, two years between the first and second stages of the PNS, undoubtedly increases the likelihood of disagreement between the measurements due to the possible change in the individuals' health conditions. Another limitation already mentioned is that the cases in which the diseases (cholesterol and CKD) may be medication-controlled

were not considered in the analyses. Likely, cases of people who declared they had received a medical diagnosis but whose condition was medication-controlled were considered false-positives.

### **Conclusion**

The study pointed out the low sensitivity of self-reported morbidity to detect diseases, especially for chronic kidney disease and altered cholesterol. Self-reported measurements can

underestimate diseases' prevalence, showing that surveys using only self-reported measurements should be analyzed with some caution. Thus, given the feasibility, laboratory components should be routinely used in population surveys to estimate the prevalence of diseases in the Brazilian population. This study's outcomes indicate that the use of health services was an essential determinant of the population's ability to correctly inform their health condition, contributing to proper disease monitoring and care.

## Collaborations

PC Pinheiro, MBA Barros, CL Szwarcwald, IE Machado, DC Malta conceived and designed the study. PC Pinheiro developed the management, exploration, and analysis of the data, the elaboration and interpretation of the results, and the discussion. All authors critically reviewed the manuscript and contributed to the entire process. All authors read, contributed to, and approved the final manuscript.

## References

1. Moreira JPL, Almeida RMVR, Luiz RR. Correção da prevalência autorreferida em estudos epidemiológicos com grandes amostras. *Cad Saude Publica* 2016; 32(12):e00050816.
2. Johnston DW, Propper C, Shields MA. Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *J Health Econ* 2009; 28(3):540-552.
3. Davillas A, Jones AM, Benzeval M. The Income-Health Gradient: Evidence From Self-Reported Health and Biomarkers in Understanding Society. In: *Panel Data Econometrics: Empirical Applications*. Amsterdam: Elsevier; 2019. p. 709-741.
4. Velakkal S, Subramanian SV, Millet C, Basu S, Stucker D, Ebrahim S. Socioeconomic Inequalities in Non-Communicable Diseases Prevalence in India: Disparities between Self- Reported Diagnoses and Standardized Measures. *PLoS One* 2013; 8(7):e68219.
5. Malta DC, Duncan BB, Schmidt MI, Machado IE, Silva AG, Bernal RTI, Pereira CA, Damacena GN, Stopa SR, Rosenfeld, LG, Szwarcwald CL. Prevalência de diabetes mellitus determinada pela hemoglobina glicada na população adulta brasileira, Pesquisa Nacional de Saúde. *Rev Bras Epidemiol* 2019; 22(Supl. 2):E190006.
6. Malta DC, Machado IE, Pereira CA, Figueiredo AW, Aguiar LK, Almeida WS, Souza MFM, Rosenfeld LG, Szwarcwald CL. Avaliação da função renal na população adulta brasileira, segundo critérios laboratoriais da Pesquisa Nacional de Saúde. *Rev Bras Epidemiol* 2019; 22(Supl. 2):E190010.
7. Malta DC, Szwarcwald CL, Machado IE, Pereira CA, Figueiredo AW, ACMGN Sá, Velasquez-Melendez G, Santo FM, PB Souza Junior, Stopa SR, Rosenfeld LG. Prevalência de colesterol total e frações alterados na população adulta brasileira: Pesquisa Nacional de Saúde. *Rev Bras Epidemiol* 2019; 22(Supl. 2):E190005.
8. Barros MBA, Francisco PMSB, Zanchetta LM, César CLC. Tendências das desigualdades sociais e demográficas na prevalência de doenças crônicas no Brasil, PNAD: 2003- 2008. *Cien Saude Colet* 2011; 16(9):3755-3768.
9. Skkiner K, Miller DR, Lincoln E, Lee A, Kazis LE. Concordance Between Respondent Self-reports and Medical Records for Chronic Conditions Experience From the Veterans Health Study. *J Ambulatory Care Manage* 2005; 28(2):102-110.
10. Okura Y, Urban LH, Mahoney DW, Jacobsen SJ, Rodeheffer RJ. Agreement between self-reported questionnaires and medical record data was substantial for diabetes, hypertension, myocardial infarction and stroke but not for heart failure. *J Clin Epidemiol* 2004; 57(10):1096-1103.
11. Szwarcwald CL, Malta DC, Souza-Júnior PRB, Almeida WS, Damacena GN, Pereira CA, Rosenfeld LG. Exames laboratoriais da Pesquisa Nacional de Saúde: metodologia de amostragem, coleta e análise dos dados. *Rev Bras Epidemiol* 2019; 22(Supl. 2):E190004.

12. Mill JG, Malta DC, Machado IE, Pate A, Pereira CA, Jaime PA, Szwarcwald CL, Rosenfeld LG. Estimativa do consumo de sal pela população brasileira: resultado da Pesquisa Nacional de Saúde 2013. *Rev Bras Epidemiol* 2019; 22(Supl. 2):E190009.
13. Souza-Júnior PR, Freitas MPS, Antonaci GA, Szwarcwald CL. Desenho da amostra da Pesquisa Nacional de Saúde 2013. *Epidemiol Serv Saúde* 2015; 24(2):207-216.
14. Kidney Disease Improving Global Outcomes. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int Suppl* 2013 [acessado 2020 Ago 6]; 3(1):1-150. Disponível em: [https://kdigo.org/wp-content/uploads/2017/02/KDIGO\\_2012\\_CKD\\_GL.pdf](https://kdigo.org/wp-content/uploads/2017/02/KDIGO_2012_CKD_GL.pdf)
15. Faludi AA, Izar MCO, Saraiva JFK, Chacra APM, Bianco HT, Afíune Neto A, Bertolami A, Pereira AC, Lottenberg AM, Sposito AC, Chagas ACP, Casella-Filho A, Simão AF, Alencar Filho AC, Caramelli B, Magalhães CC, Magnoni D, Negrão CE, Ferreira CES, Scherr C, Feio CMA, Kovacs C, Araújo DB, Calderaro D, Gualandro DM, Mello Junior EP, Alexandre ERG, Sato IE, Moriguchi EH, Rached FH, Santos FC, Cesena FHY, Fonseca FAH, Fonseca HAR, Xavier HT, Pimentel IC, Giuliano ICB, Issa JS, Diament J, Pesquero JB, Santos JE, Faria Neto JR, Melo Filho JX, Kato JT, Torres KP, Bertolami MC, Assad MHV, Miname MH, Scartezini M, Forti NA, Coelho OR, Maranhão RC, Santos Filho RD, Alves RJ, Cassani RL, Betti RTB, Carvalho T, Martinez TLR, Giraldez VZR, Salgado Filho W. Atualização da Diretriz Brasileira de Dislipidemias e Prevenção da Aterosclerose – 2017. *Arq. Bras. Cardiol* 2017 Aug [cited 2020 Aug 7]; 109(2 Supl. 1):1-76. Available from: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0066-782X2017001100001&lng=en](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0066-782X2017001100001&lng=en)
16. Ning M, Zhang Q, Yang M. Comparison of self-reported and biomedical data on hypertension and diabetes: findings from the China Health and Retirement Longitudinal Study (CHARLS). *BMJ Open* 2016; 6:e009836.
17. Crimmins EM, Vasunilashorn SM. Biodemography: Adding Biological Insight into Social, Economic, and Psychological Models of Population and Individual Health Change with Age. In: Kenneth F, Ferraro LKG, editors. *George Handbook of Aging and the Social Sciences*. Amsterdam: Elsevier; 2016. p. 55-75.
18. Harris KM, Schorpp KM. Integrating Biomarkers in Social Stratification and Health Research. *Annu Rev Sociol* 2018; 44:361-386.
19. Lima-Costa ME, Peixoto SV, Firmo JOA, Uchoa E. Validade do diabetes auto-referido e seus determinantes: evidências do projeto Bambuí. *Rev Saude Publica* 2007; 41(6):947-953.
20. Wu SC, Li CY, Ke DS. The agreement between self-reporting and clinical diagnosis for selected medical conditions among the elderly in Taiwan. *Public Health* 2000; 114(2):137-142.
21. Molenaar EA, Van Ameijden EJ, Grobbee DE, Numans ME. Comparison of routine care self-reported and biometrical data on hypertension and diabetes: results of the Utrecht Health Project. *Eur J Public Health* 2007; 17(2):199-205.
22. Goldman N, Lin I-F, Weinstein M, Lin Y-H. Evaluating the quality of self-reports of hypertension and diabetes. *J Clin Epidemiol* 2003; 56(2):148-154.
23. Stopa SR, Malta DC, Monteiro CN, Szwarcwald CL, Goldbaum M, Galvão CCL. Acesso e uso de serviços de saúde pela população brasileira, Pesquisa Nacional de Saúde 2013. *Rev Saude Publica* 2017; 51(Supl. 1):3s.
24. Chrestani MAD, Santos IS, Matijasevich AM. Hipertensão arterial sistêmica auto-referida: validação diagnóstica em estudo de base populacional. *Cad Saude Publica* 2009; 25(11):2395-2406.
25. Vargas CM, Burt VL, Gillum R, Pamuk ER. Validity of self-reported hypertension in the National Health and Nutrition Examination Survey III, 1988-1991. *Prev Med* 1997;26(5 Pt 1):678-685.
26. Lee HY. Socioeconomic Disparities in the Prevalence, Diagnosis, and Control of Hypertension in the Context of a Universal Health Insurance System. *J Korean Med Sci* 2017; 32(4):561-567.
27. Instituto Brasileiro de Geografia e Estatística (IBGE). *Síntese de indicadores sociais: uma análise das condições de vida da população brasileira*. Rio de Janeiro: IBGE; 2016.
28. Onur I, Velamuri M. The gap between self-reported and objective measures of disease status in India. *PLoS One* 2018; 13(8):e0202786.

Article submitted 02/12/2020

Approved 02/12/2020

Final version submitted 04/12/2020

Chief editors: Maria Cecília de Souza Minayo, Romeu Gomes, Antônio Augusto Moura da Silva

