

Nota Técnica

Nº 59

Diest

Diretoria de Estudos e Políticas do
Estado, das Instituições e da Democracia

Novembro de 2022

IMPUTAÇÃO DA INFORMAÇÃO DE RAÇA/ COR NA RAIS PARA O SETOR PÚBLICO BRASILEIRO

Leonardo Silveira



Governo Federal

Ministério da Economia

Ministro Paulo Guedes

ipea

Instituto de Pesquisa
Econômica Aplicada

Fundação pública vinculada ao Ministério da Economia, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidente

Erik Alencar de Figueiredo

Diretor de Desenvolvimento Institucional

André Sampaio Zuvanov

Diretor de Estudos e Políticas do Estado, das Instituições e da Democracia

Flavio Lyrio Carneiro

Diretor de Estudos e Políticas Macroeconômicas

Marco Antônio Freitas de Hollanda Cavalcanti

Diretor de Estudos e Políticas Regionais, Urbanas e Ambientais

Nilo Luiz Saccaro Junior

Diretor de Estudos e Políticas Setoriais, de Inovação, Regulação e Infraestrutura

João Maria de Oliveira

Diretor de Estudos e Políticas Sociais

Herton Ellery Araújo

Diretor de Estudos Internacionais

Paulo de Andrade Jacinto

Coordenador-Geral de Imprensa e Comunicação Social (substituto)

João Cláudio Garcia Rodrigues Lima

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>

© Instituto de Pesquisa Econômica Aplicada – ipea 2022

EQUIPE TÉCNICA

Leonardo Silveira

Pesquisador associado à Diretoria de Estudos e Políticas do Estado, das Instituições e da Democracia do Instituto de Pesquisa Econômica Aplicada (Diest/Ipea). *E-mail*: <leonardo.silveira@ipea.gov.br>.

DOI: <http://dx.doi.org/10.38116/ntdiretoria59>

As publicações do Ipea estão disponíveis para *download* gratuito nos formatos PDF (todas) e EPUB (livros e periódicos). Acesse: <<http://www.ipea.gov.br/portal/publicacoes>>.

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério da Economia.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte.
Reproduções para fins comerciais são proibidas.

SUMÁRIO

1 INTRODUÇÃO	5
2 A RAIS PARA O ESTUDO DO SETOR PÚBLICO: REPRESENTATIVIDADE ESTATÍSTICA E AUSÊNCIA DE DADOS RACIAIS	5
3 MÚLTIPLOS OLHARES SOBRE A RAÇA	7
4 ESTRATÉGIAS DE IMPUTAÇÃO DOS DADOS FALTANTES	9
5 COMPOSIÇÃO E DIFERENCIAL DE RENDA RACIAL	13
6 CONSIDERAÇÕES E RECOMENDAÇÕES DE USO	19
REFERÊNCIAS	20

1 INTRODUÇÃO

A Relação Anual de Informações Sociais (Rais) é uma fonte de dados quase censitária sobre o mercado formal brasileiro coletada desde 1975. Para o setor público, em especial, é utilizada desde, pelo menos, a década de 1980, dado que ela permite que recortes por esferas, poderes e entes subnacionais continuem estatisticamente representativos, ao contrário de outras pesquisas amostrais. Contudo, em relação à variável raça/cor, a Rais não permite análises detalhadas no setor público, devido ao grande número de casos não declarados – em 2019, esse percentual é de 100% entre os estatutários. Desse modo, esta nota propõe uma estratégia para contornar essa falta de dados raciais na Rais, mediante registros prévios dos trabalhadores e imputação por meio de probabilidade. Além disso, esse processo é descrito e são realizadas comparações entre a Rais, o Sistema Integrado de Administração de Pessoal (Siape) e a Pesquisa Nacional por Amostra de Domicílios (PNAD) Contínua.

Para além do alto percentual de não declaração da variável raça/cor na Rais para o setor público, ela suscita outra discussão. Nela, o formato classificatório é a heteroclassificação, isto é, a declaração racial é feita pelo empregador ou por algum funcionário administrativo responsável. Esse formato possui contrastes em relação à autodeclaração, mais usual, e leva a resultados semelhantes, mas nem sempre sobrepostos – conforme observado por Paixão *et al.* (2012), por exemplo.

Sendo assim, este trabalho: i) explica as características da Rais que permitem a análise do corpo de funcionários públicos no Brasil; ii) analisa como o alto percentual de não declaração da raça interfere nas suas estimativas; e iii) propõe uma solução para tal problema. Para discutir tais limitações bem como contextualizar a Rais e as formas de classificação racial, o texto é dividido em outras cinco partes, além desta introdução. Na segunda seção, apresenta-se a Rais e apontam-se seus usos, seus potenciais e suas limitações, especialmente na análise do setor público. Na terceira seção, são expostos os contrastes e os fundamentos das diferentes formas de classificação racial. As estratégias de imputação das categorias raciais são destrinchadas na quarta seção, de modo a indicar os caminhos adotados. A quinta seção compara os resultados obtidos por essas estratégias com os dados do Siape e da PNAD Contínua – no primeiro caso, estão os indivíduos vinculados caso a caso, o que permite estimar a consistência das imputações. Por último, conclui-se com a apresentação de recomendações e limitações no uso da nova informação racial gerada, de modo a indicar também apontamentos para novas explorações.

2 A RAIS PARA O ESTUDO DO SETOR PÚBLICO: REPRESENTATIVIDADE ESTATÍSTICA E AUSÊNCIA DE DADOS RACIAIS

A Rais é uma fonte única de dados sobre o mercado de trabalho formal no Brasil. Coletada desde 1975, ela serviu tanto como base de controle e gestão de políticas públicas quanto como fonte de dados para estudos sobre o emprego e o salário públicos (Sanson e Moutinho, 1987; Lopez e Guedes, 2020; Vaz e Nakamura, 2020; Nunes e Teles, 2021; Silveira *et al.*, 2021; Santos *et al.*, 2018) e sobre desigualdades raciais (Silveira, 2019; Paixão *et al.*, 2012; Cornwell, Rivera e Schmutte, 2017; Noronha e Vilela, 2020; Almeida, 2019; Dias e Rocha, 2021; Miller e Schmutte, 2021), além de outros usos que extrapolam esses temas.

De maneira geral, os pontos favoráveis ao uso da Rais estão relacionados a três características principais: sua representatividade estatística, seu formato em painel e as informações contidas na base. Em termos de representatividade, a Rais possui uma cobertura *quase censitária* do mercado de trabalho formal brasileiro. Devido a constrangimentos e incentivos do governo às empresas, o estabelecimento empregador deve registrar todos os vínculos funcionais em um ano corrente, incluindo aqueles que foram desligados. Isso faz que a Rais cubra todo o setor formal, inclusive o setor público, e sua cobertura seja *quase censitária*, devido ao fato de o empregador nem sempre estar incluído entre os vínculos¹ e à existência de eventuais erros de preenchimento.

Em geral, as informações raciais estão presentes em diferentes pesquisas domiciliares e registros administrativos no Brasil, sendo objeto de profundas análises para entender seus potenciais e suas limitações (Osório, 2003; Petrucci, 2013; Senkevics, Machado e Oliveira, 2016). Em comparação às bases usualmente analisadas, como o Censo, a PNAD e a PNAD Contínua, ambas do Instituto Brasileiro de Geografia e Estatística (IBGE), a Rais possui uma representatividade maior para recortes regionais, uma vez que se trata do *quase* universo do setor formal. Por exemplo, a PNAD Contínua é representativa até o nível estadual e para regiões metropolitanas, ao passo que na Rais um recorte municipal, ou até mesmo intramunicipal, é representativo, quando se trata do setor formal. Para além da representatividade geográfica, outros recortes são representativos por meio da Rais, especialmente no caso da

1. Sobre isso, ver o trabalho de Dias e Rocha (2021) para encontrar informações raciais sobre o empregador.

análise para o setor público. Em pesquisas domiciliares amostrais, a representatividade de funcionários públicos fica comprometida, e não permite que sejam feitas comparações entre trabalhadores de diferentes níveis de poder e esferas federativas, enquanto na Rais se tem o universo de vínculos no Brasil. Isso permite que sejam feitas análises para grupos menores, como vínculos de municípios ou estados específicos, ou ainda realizar amostragem de grupos maiores, como os trabalhadores do Poder Executivo municipal no Brasil, que conta com mais de 6 milhões de vínculos.²

O segundo ponto favorável da Rais diz respeito à possibilidade de uso longitudinal em painel das informações. Isto é, em divulgações identificadas se há informações dos vínculos em diferentes anos durante a trajetória profissional do indivíduo, se ele estiver no mercado formal. Bases com tal estrutura são raras, dados os custos de realização e manutenção de pesquisas e registros desse porte. A Rais, por sua vez, além da sua representatividade quase censitária, permite acompanhar as trajetórias profissionais, de modo a analisar aspectos de mobilidade ocupacional intrageracional, salarial, regional, entre outras. Ao usar dados de painel em modelos estatísticos, obtêm-se pressupostos menos rígidos, por controlar variáveis não observáveis ao longo do tempo, seja do vínculo, seja do empregador (como utilizado por Saperstein e Penner, 2012).

Outro ponto favorável à utilização da Rais são as informações contidas. Nela, encontram-se variáveis como o tipo de contrato estabelecido, o tamanho do estabelecimento empregador, a ocorrência e os motivos de afastamento do indivíduo, a nacionalidade e se possui alguma deficiência. Em sua disponibilização mais restrita, a Rais possui o PIS do trabalhador, o que permite que a trajetória de empregos seja reconstruída de maneira retroativa. Desse modo, embora coletada de maneira transversal a cada ano, a base pode se tornar um painel longitudinal, em que cada trabalhador pode ser acompanhado ao longo do tempo.

No entanto, é necessário jogar luz sobre algumas limitações da Rais, principalmente em comparação às pesquisas domiciliares amostrais, sendo: i) a falta de cobertura da economia informal; ii) a falta de informações para períodos fora do mercado de trabalho; iii) a escassez de informações sociodemográficas; e iv) a não uniformidade no treinamento de preenchimento das informações.

A primeira delas é o fato de não conter informações sobre trabalhadores informais, o que não permite análises para esse grupo, além de gerar um viés de seleção que não é mensurado. Isto é, as dinâmicas no mercado formal são influenciadas pelas dinâmicas do mercado informal, as quais ficam de fora do escopo da Rais.³ Isso leva a um segundo problema, que é o fato de a análise longitudinal ser possível apenas quando o trabalhador se encontra no mercado formal, não havendo informações do período fora do mercado, em situação de desemprego ou na informalidade.

O terceiro problema apontado para a Rais é a escassez de informações sociodemográficas, o que a torna restrita ao gênero, à raça e à idade do trabalhador, e não possui informações familiares/domiciliares e outras, que são encontradas em pesquisas domiciliares amostrais. Tais informações influenciam em resultados no mercado de trabalho, mas ficam de fora das análises com base na Rais.

Por último, aponta-se a falta de uniformidade no treinamento dos responsáveis pelo preenchimento da base. Em pesquisas domiciliares amostrais, o número de entrevistadores é menor, e esses recebem treinamentos que homogeneizam as respostas. A Rais, por sua vez, possui manuais de preenchimento, e grande parte das variáveis dizem respeito a aspectos burocráticos, como o tipo de contrato e o número de funcionários do estabelecimento empregador. Não é incomum, contudo, serem encontrados erros de preenchimento, especialmente nas variáveis natureza jurídica, escolaridade e raça/cor do trabalhador. Para lidar com o primeiro caso, ausência de informações sobre trabalhadores informais, o Atlas do Estado Brasileiro⁴ e Santos *et al.* (2018) propuseram formas de contornar tal problema.

Já no caso da raça, outras questões se colocam e serão discutidas a seguir. É importante destacar que muitos estudos com base na Rais foram realizados para o setor público brasileiro (Sansón e Moutinho, 1987; López e Guedes, 2020; Vaz e Nakamura, 2020) e para a desigualdade racial no mercado de trabalho (Paixão *et al.*, 2012; Cornwell, Riviera e Schmutte, 2017; Almeida, 2019; Noronha e Vilela, 2020; Miller e Schmutte, 2021). Contudo, essas duas análises não andam juntas, já que o número de não respostas para a variável racial é superior a 90% no setor público, enquanto no setor privado ela gira em torno de 20%. A proposta de contornar as não respostas visa, portanto, permitir que análises sobre desigualdades raciais sejam realizadas com base na Rais para o setor público.

2. Ver Silveira (livro).

3. Almeida (2019) utilizou variáveis do Censo Demográfico para estimar o viés de seletividade da Rais em seus modelos para análise da renda.

4. Disponível em: <<https://bit.ly/3ys9tXa>>.

A classificação racial é mensurada pela Rais e foi utilizada em diferentes estudos (Silveira, 2019; Paixão *et al.*, 2012; Cornwell, Riviera e Schmutte, 2017; Noronha e Vilela, 2020; Almeida, 2019; Dias e Rocha, 2021). A sua forma de classificação, porém, é distinta da autoclassificação, mais usual em pesquisas amostrais realizadas pelo IBGE. Na Rais, a raça é classificada por um observador externo, seja o empregador, seja algum funcionário responsável – forma de classificação chamada na literatura de heteroclassificação, alterclassificação ou classificação por um observador externo. Isso traz vantagens e desvantagens para o estudo das desigualdades raciais no setor público, e no mercado de trabalho, de maneira geral⁵ – o que será explorado na seção seguinte.⁶

Por ora, demarca-se que se tem na Rais uma base com representatividade e informações únicas para o estudo do mercado de trabalho, dos salários e do funcionalismo público no Brasil. Por meio dela, podem-se fazer recortes geográficos, por setores da economia, seguindo as trajetórias de indivíduos e de perfis ocupacionais ao longo do tempo. No caso da raça, podem-se fazer estimativas para os diferentes grupos e analisar a trajetória de indivíduos desde a sua entrada no mercado formal. O alto percentual de não resposta para tal variável, contudo, especialmente no setor público, deixa uma lacuna nos estudos que analisam esses dois aspectos. Na seção seguinte, são analisadas as implicações da heteroclassificação para o estudo do funcionalismo público e das desigualdades raciais no Brasil bem como as estratégias adotadas para contornar tais problemas na Rais.

3 MÚLTIPLOS OLHARES SOBRE A RAÇA

A raça é uma forma de categorização de grupos e indivíduos construída em interação social. Por meio dela, grupos orientam suas ações em relação a atitudes políticas, formas de tratamento, preferências por relações sociais de amizade, trabalho, entre outros. Contudo, a identificação e a classificação racial são fluidas, já que não são biologicamente determinadas, e desse modo, passíveis de variações ao longo do tempo, do contexto e das formas de mensuração adotadas.

Ao longo do tempo, categorias se tornam mais abrangentes, ou refletem uma compreensão racial distinta. Um exemplo é o aumento da classificação na categoria preta desde os anos 2000 no Brasil, dado um novo reconhecimento por parte da população (Monk, 2013; Jesus e Hoffman, 2020; Senkevics, 2022). Por contexto, a compreensão racial varia, uma vez que categorias que são usuais em um lugar, não o são em outro, de maneira que isso ocorre, por exemplo, em diferentes países (Telles e Paschel, 2014), ou regiões (Monk, 2013; Silveira, 2019). Por fim, pode-se apontar como as categorias (quais são os termos utilizados, ou se as respostas são abertas ou fechadas) direcionam as respostas, como o caso da categoria “morena” no Brasil, popularmente utilizada, mas não adotada por órgãos oficiais (Silva, 1999b; Telles, 1995).

Os exemplos de tais variações são vastos e amplamente analisados na literatura das ciências sociais.⁷ Aqui, seguem-se os apontamentos de Roth (2010), que defende que cada forma de coletar a informação racial reflete uma dimensão desse construto social. Seja em pesquisas etnográficas, seja em *surveys*, ou ainda em registros administrativos, a classificação pode ser feita pelo entrevistado ou pelo entrevistador, por categorias abertas ou pré-determinadas, baseadas em aparência ou em interação. Com isso, Roth (*op. cit.*) identifica sete formas de classificação racial em pesquisas nas ciências sociais. O “problema”, contudo, é que nem sempre essas formas coletam a mesma categoria, já que a raça não é uma categoria biológica nem estática, sendo elaborada e reelaborada em interação. Além disso, a autora aponta que tais inconsistências classificatórias são importantes para demonstrar que a raça é uma construção coletiva, o que diminuiria as chances de reificação das categorias raciais (*op. cit.*).

Entre as sete formas de classificação trazidas por Roth (2010), destacam-se duas principais para se pensar a variável raça na Rais, a expressa e a observada. A primeira diz respeito à autoidentificação com opções em categorias fechadas, ou seja, conforme as pesquisas do IBGE e grande parte de questionários e *surveys* no mundo. Essa forma é usual por respeitar a autodeterminação dos grupos, conforme cancelado pela Conferência de Durban, em 2001. Como sustenta Saperstein (2012), a autoclassificação mensura aspectos sobre a trajetória de indivíduos relacionados à ancestralidade e a tratamentos recebidos na vida social.

A classificação observada, por seu turno, trata da escolha de um observador externo, a partir de categorias pré-definidas. Ou seja, a classificação racial feita na Rais segue esse direcionamento, além de ser influenciada por dois aspectos destacados por Roth (2010), a interação e a aparência. No primeiro caso, o classificador opta por uma categoria após a convivência e a interação com o indivíduo, enquanto no segundo caso ele utiliza aspectos estéticos que

5. Consultar Roth (2010) e Silveira (2019).

6. Para além da falta de uniformidade do treinamento no caso de coleta, é notado que o baixo percentual de respostas da variável cor/raça na Rais para o setor público está vinculado ao fato do seu preenchimento não ser obrigatório.

7. Ignatiev (1995); Carvalho, Wood e Andrade (2004); Miranda (2015); Silveira e Tomás (2019); de Micheli (2019); Loveman e Muniz (2007).

o classificador julga como determinantes para a classificação racial. Na Rais, o classificador pode tanto ter convivido com o trabalhador a ser classificado quanto apenas classificar segundo uma fotografia. A despeito das diferenças para a autoclassificação, Roth (*op. cit.*) sustenta que a heteroclassificação possui vantagens para o estudo da discriminação, já que se trata de uma aproximação de como aquele indivíduo é percebido racialmente em interações sociais.

Desse modo, o uso de formas de classificação distintas, como nas pesquisas domiciliares do IBGE e na Rais, traz implicações para os resultados. Ambas dialogam sobre percepções e construções coletivas sobre a raça, mas a autoclassificação reflete trajetórias individuais e de grupos, enquanto a heteroclassificação está mais relacionada à interação mais imediata e a estereótipos raciais.

Mais especificamente quanto ao uso da Rais, Paixão *et al.* (2012) apontam as implicações da heteroclassificação em relação à autoclassificação na PNAD. Segundo os autores, ao comparar as duas fontes de dados, chega-se às conclusões descritas a seguir.

- 1) Ambas “produzem informações distintas, porém complementares, e igualmente úteis para o entendimento da dinâmica do mercado de trabalho do Brasil” (Paixão *et al.*, 2012, p. 23).
- 2) Na Rais, há a possibilidade de não declarar a raça, enquanto na PNAD isso não ocorre.
- 3) Na Rais, há uma tendência de “branquear” as classificações, de maneira que nessa há 2 milhões a mais de brancos se comparada com a PNAD.
- 4) Esse “branqueamento” é maior nas regiões Sul e Sudeste.
- 5) Apesar dessas questões, a composição racial da população celetista era similar nas duas bases.
- 6) O rendimento salarial médio é sobrestimado na Rais para todos os grupos, mas o diferencial salarial entre eles é coerente com a PNAD.

De maneira geral, portanto, o que as conclusões de Paixão *et al.* (2012) indicam é que as classificações nas duas bases são distintas e não serão totalmente intercambiáveis, mas, ao mesmo tempo, elas não podem ser desprezadas como fontes de informação pública. Além disso, a Rais não cria viés sensível nas principais informações extraídas para os estudos de desigualdade racial, que são as disparidades socioeconômicas. O uso dessa variável é, portanto, endossado, embora sejam necessárias análises, tratamento da informação disponível e aprimoramento na coleta.⁸

Desse modo, é importante levar em consideração que, usualmente, estimativas baseadas em heteroclassificação levam a resultados distintos daqueles baseados em autoclassificação. No Brasil, a heteroclassificação está associada a outras variáveis, como escolaridade, renda e região – por exemplo, o entrevistador tende a classificar como “brancas” pessoas com maior escolaridade (Silveira e Tomás, 2019); assim como a localização nas regiões Sul e Sudeste tende a extremos branco/preto, enquanto no Nordeste a categoria parda é a mais utilizada (Monk, 2013; Silveira, 2019). Já a autoclassificação é balizada por outros direcionamentos, como a maior escolaridade, que é associada às categorias “extremas” branca e preta, o que seria um indicador de maior “consciência racial” (Telles e Paschel, 2014).

Nesse contexto, portanto, ao mesmo tempo que é esperado que a auto e a heteroclassificação tenham aproximações e que a maioria das classificações concorde entre si, também é esperado que possam existir diferenças relevantes em alguns subgrupos. Analisar a heteroclassificação diz respeito a aspectos de interação social contextual, ao passo que a autoclassificação reflete trajetórias de indivíduos e grupos, e ambas podem variar ao longo do tempo, devido a mudanças em aspectos culturais e episódios enfrentados pelos indivíduos.

Nesse sentido, a Rais surge como uma fonte rica de dados, a qual cobre aspectos longitudinais; informações relativas às dinâmicas de entrada e saída do mercado de trabalho; e a representatividade estatística de grupos que não são cobertos pelas pesquisas domiciliares amostrais. Por outro lado, o uso da variável raça traz ressalvas relacionadas à heteroclassificação e ao alto percentual de não declaração no setor público. O esforço aqui retratado é o de transformar a variável raça da Rais factível de ser utilizada inclusive para este setor, visando incrementar a produção de dados sobre inserção, permanência, promoção/mobilidade e desigualdades entre os funcionários públicos nas três esferas e nos três poderes no Brasil.

8. Diferenças entre auto e heteroclassificação podem ser encontradas, por exemplo, em Roth (2010), Saperstein (2012) e Vargas (2014). No Brasil são encontradas em obras como as de Silva (1999b); Muniz (2012); Bailey, Loveman e Muniz (2013); Ribeiro (2017); e Silveira e Tomás (2019).

A variável raça na Rais é preenchida a cada ano para cada trabalhador, e tem como opções as categorias *indígena, branca, preta, parda, amarela e não informada*. Esse preenchimento é feito por algum responsável do estabelecimento empregador, que pode ser o próprio empregador ou um empregado responsável pelo setor administrativo ou pelos recursos humanos. O Ministério da Economia disponibiliza um manual de preenchimento, que, porém, não é específico com relação à variável cor/raça, além de o não preenchimento não ser um impedimento para submeter registros de vínculos.

Considerando apenas os trabalhadores em regime estatutário, nenhum deles tem a raça classificada no ano de 2019.⁹ Em comparação, no setor privado, apenas 13,2% não têm a raça preenchida para este mesmo ano. Tem-se, portanto, que a raça não é devidamente declarada no setor público brasileiro, por isso não permite fazer inferências sobre o tema – mesmo considerando outros regimes que não seja o estatutário.¹⁰ Entretanto, deve-se levar em consideração que muitos dos vínculos do setor público já passaram pelo setor privado no Brasil, sendo mais provável de terem sido classificados em relação à raça – em grande parte, em mais de uma ocasião. Isso permite que se tenha alguma informação sobre como este trabalhador é visto racialmente por empregadores e outros trabalhadores, e qual foi a categoria mais recorrente. Essa estratégia já foi adotada por Noronha (2018), em análise sobre imigrantes internacionais, e por Silveira (2019), que analisou a variação das desigualdades raciais ao se adotar a primeira classificação racial na Rais e a raça em que a classificação foi mais recorrente para cada indivíduo.¹¹ Desse modo, a primeira estratégia, de utilizar as classificações anteriores, aumenta o percentual de dados raciais preenchidos para os servidores estatutários de 0% para 73,8% no ano de 2019.

4.1 Primeira estratégia

A primeira estratégia utilizada foi a identificação das classificações anteriores e assumir a moda (a classificação mais recorrente ao longo dos anos) da variável para aquele indivíduo para um conjunto de anos de exposição, ou seja, que aparece na Rais. Isso acontece pois muitos vínculos empregatícios tiveram a raça classificada em algum ponto do tempo. Contudo, essa classificação pode ser alterada de acordo com a percepção do observador, por isso utiliza-se a moda como um recurso para mensurar em qual categoria racial aquele indivíduo é mais comumente percebido pelas demais pessoas.

TABELA 1

Proporção de não respostas na variável raça/cor por poder e esfera dos trabalhadores estatutários – Brasil (2019)
(Em %)

Esfera	Poder	Original	Após a primeira estratégia
Estadual	Executivo	100,0	26,0
Estadual	Judiciário	100,0	46,8
Estadual	Legislativo	100,0	31,1
Federal	Executivo	100,0	35,6
Federal	Judiciário	100,0	35,8
Federal	Legislativo	100,0	64,9
Municipal	Executivo	100,0	22,6
Municipal	Legislativo	100,0	24,9
Total	-	100,0	26,2

Fontes: Atlas do Estado Brasileiro (2022); e Rais (2019).

Os resultados da tabela 1 demonstram que a primeira estratégia reduz o percentual de dados raciais faltantes de 100,0%, entre os estatutários, para 26,2%. Isso quer dizer que, em algum momento de suas trajetórias, 73,8% tiveram sua raça registrada pelo menos uma vez na Rais. Observa-se na tabela 1 que o percentual cai, principalmente no Executivo municipal, para 22,6%, enquanto no Legislativo federal esse percentual continua em torno de 64,9%.

9. Considerando os outros regimes contratuais, o percentual de preenchimento é de 6,8% (Executivo estadual), 0,3% (Judiciário estadual), 0,1% (Legislativo estadual), 6,5% (Executivo federal), 0,1% (Judiciário federal), 0,0% (Legislativo federal), 9,4% (Executivo municipal) e 3,0% (Legislativo municipal).

10. O preenchimento da variável cor/raça não é obrigatório. Informações coletadas informalmente com setores de pessoal da administração pública indicam que parte da rotina de preenchimento do cadastro sugere não preencher tal campo. Contudo, cabem estudos qualitativos, tanto no setor público quanto no privado, para compreender melhor como a raça e outras variáveis são registradas.

11. Os resultados indicaram que há uma tendência de a classificação racial ser associada ao *status* socioeconômico, mas não há variações significativas nos indicadores de desigualdade – raça, índice socioeconômico e disparidade escolar.

TABELA 2

Frequência da classificação racial dos trabalhadores estatutários – Brasil (2019)

Frequência	Servidores	Proporção (%)
Mais de uma categoria	1.070.868	16,7
Sempre na mesma categoria	3.656.195	57,1
Nunca foram classificados	1.677.015	26,2
Total	6.404.078	100,0

Fontes: Atlas do Estado Brasileiro (2022); e Rais (2019).

Outra informação relevante diz respeito à frequência com que a categoria racial foi utilizada anteriormente. Isto é, de 100% dos casos, 26,2% não tiveram a raça classificada em nenhum momento na Rais. Dos 73,8% que tiveram essa informação registrada, 57,1% sempre foram classificados na mesma categoria ao longo do tempo, e apenas 16,7% tiveram mais de uma categoria aferida, ou seja, foram reclassificados em algum momento.¹²

4.2 Segunda estratégia

A primeira estratégia, portanto, resolve 73,8% dos dados faltantes para a raça na Rais, entre os estatutários, no ano de 2019. Restam, entretanto, 26,2% de servidores sem classificação. A segunda estratégia permite chegar à totalidade de vínculos com classificação, baseada em imputação múltipla. Com base na estratégia de imputação *multivariate imputation by chained equations* (Mice) – Buuren e Groothuis-Oudshoorn (2011) –, os casos em que a raça era um dado faltante, ela foi imputada com base nas características individuais, regionais e setoriais. No caso do setor público brasileiro, utilizou-se: salário, gênero, escolaridade, idade, tipo de vínculo, esfera, poder e Unidade da Federação (UF).

4.2.1 O uso da imputação

Dadas as características do formato e da coleta da informação racial em registros administrativos, elas precisam ser objeto de cuidadosa análise, uma vez que haverá inconsistências e limitações para o seu uso (Paixão *et al.*, 2012; Braz *et al.*, 2013). No caso da Rais, o alto percentual de ausência de respostas é um limitador para a sua análise, mas dada a redução dos casos para menos de um quarto, torna-se possível a aplicação de técnicas de imputação. Tal estratégia já foi utilizada para casos faltantes de raça em bases de dados de mortalidade (Thompson *et al.*, 2017; Zhang *et al.*, 2022).

Os dados faltantes por não resposta a determinados itens de um questionário são bastante comuns na análise de dados por *survey* ou mesmo por registros administrativos. Para contornar essa falta, algumas sugestões são feitas, dependendo do padrão observado para essas não respostas – seja por ignorar as não respostas, seja pelo preenchimento baseado na média daquela variável, ou, por fim, pela imputação (Albieri, 1989). Se o padrão de ausência dos dados permitir, a imputação atribui a cada caso um valor, ou categoria, de acordo com outras variáveis observadas e devidamente preenchidas.

Para que essa estratégia seja levada a cabo, é necessário analisar e diagnosticar o padrão de ausência de respostas, que pode ser de três tipos. O primeiro tipo é o de ausência completamente aleatória – *missing completely at random* (MCAR) –, de maneira que nenhuma das variáveis observáveis ou não observáveis seja suficiente para prever a ausência. Trata-se de um tipo raro de padrão, normalmente estimulado por recodificações aleatórias para fins analíticos mais do que recorrentes em *surveys* e registros administrativos.

O segundo tipo, mais comum, é o de ausência aleatória, ou *missing at random* (MAR). Ela é categorizada dessa maneira se outras variáveis observadas na base de dados permitirem prever a sua ausência. Nesse tipo, os casos com ausência de resposta podem ser preenchidos por uma função de outras variáveis, chegando a um valor aproximado, sendo, portanto, um caso menos restritivo que os demais. Por último, tem-se a ausência não aleatória, ou *missing not at random* (MNAR), em que apenas um valor não observado seja possível de prever a ausência.

Nesse contexto, em casos de ausência aleatória (MAR), é usual aplicar técnicas de imputação. Nesse caso, foi utilizada a técnica Mice (Buuren e Groothuis-Oudshoorn, 2011), que estima numerosos modelos de regressão com base nas variáveis selecionadas. Segundo Azur *et al.* (2011), essas equações encadeadas são destrinchadas em quatro passos, que envolvem a substituição do valor ausente pela média observada, posteriormente atribuída como caso ausente,

12. Para se ter um parâmetro de comparação, Silveira (2019) demonstra que, entre 2008 e 2015, 9% dos casos totais da Rais tiveram a sua raça reclassificada.

seguido de uma regressão nos valores ausentes em função das demais covariáveis e, por fim, pela recodificação do valor ausente no valor mais provável segundo os modelos de regressão. Esse processo é repetido após múltiplas iterações, na qual é assumido aquele com melhor ajuste (Azur *et al.*, 2011; Hegde *et al.*, 2019).

Segundo os autores do pacote no R, “a filosofia por trás da metodologia Mice é que múltiplas imputações são melhores como uma sequência de pequenos passos, em que cada qual requer checagem de diagnósticos” (Buuren e Groothuis-Oudshoorn, 2011, p. 4). Nas seções seguintes, tais diagnósticos serão apresentados, demonstrando as estatísticas descritivas e a estimação de desigualdades com base na variável raça antes e depois da imputação, visando validar os procedimentos.

4.2.2 Diagnóstico da não declaração da raça

A tabela 3 apresenta a distribuição dos casos ausentes para os anos de 2004 e 2019, considerando todos os funcionários públicos brasileiros. Em 2004, 20,2% não tinham registro de anos anteriores, ao passo que em 2019 esse valor sobe para 24,6%. Por gênero, observa-se que, em ambos os anos, os homens têm uma pequena diferença em relação às mulheres, com mais ausências. Já em relação à escolaridade, não há um padrão definido por ano.

Quando se refere às variáveis institucionais do setor público, tem-se que a esfera municipal possui menos ausências que as demais. Isso não é necessariamente relacionado à essa esfera, mas aos funcionários que ocupam cargos em sua administração. Isto é, é mais provável que esses trabalhadores já tenham ocupado funções no setor privado e tenham tido a raça classificada em outras oportunidades. Ao contrário, quando se observam os níveis de poder, o Judiciário é aquele com maior percentual de ausências. Considerando as esferas e os poderes, uma hipótese que pode estar relacionada à falta da classificação racial prévia é o fato de se tratar de funcionários que têm na esfera federal e no Poder Judiciário o seu primeiro emprego, não tendo sido registrado no setor privado anteriormente.

TABELA 3

Estatísticas descritivas de declaração racial dos funcionários públicos no Brasil após a primeira estratégia de preenchimento de dados faltantes (2004-2019)

Variáveis	Categorias	2004			2019		
		Declarado	Não declarado	Total	Declarado	Não declarado	Total
Gênero	Masculino	78,7	21,3	100,0	74,8	25,2	100,0
	Feminino	80,5	19,5	100,0	75,8	24,2	100,0
Escolaridade	Até fundamental incompleto	82,9	17,1	100,0	76,2	23,8	100,0
	Fundamental completo ou médio incompleto	79,9	20,1	100,0	70,6	29,4	100,0
	Médio completo ou superior incompleto	79,9	20,1	100,0	72,7	27,3	100,0
	Superior completo	78,5	21,5	100,0	78,1	21,9	100,0
Esfera	Estadual	69,9	30,1	100,0	73,7	26,3	100,0
	Federal	65,2	34,8	100,0	65,9	34,1	100,0
	Municipal	90,3	9,7	100,0	77,7	22,3	100,0
Poder	Executivo	80,5	19,5	100,0	76,4	23,7	100,0
	Judiciário	57,2	42,8	100,0	56,0	44,0	100,0
	Legislativo	78,9	21,1	100,0	67,5	32,5	100,0
Idade média		40,05***	43,59***	40,77	44,08***	42,39***	43,66
Renda média		2590,61***	3842,61***	2.843,92	4256,73***	5045,77***	4.450,85
Total		79,8	20,2	100,0	75,4	24,6	100,0

Fontes: Atlas do Estado Brasileiro (2022); e Rais (2004; 2019).

De maneira geral, há pequenas variações entre os anos de 2004 e de 2019. Pode-se notar ainda que a renda média dos casos ausentes é mais alta do que aqueles com preenchimento, estando isso, porém, relacionado com o maior número de casos na esfera federal e no Poder Judiciário, que usualmente possuem maiores salários.

Já na tabela 4, tem-se o percentual de casos ausentes por UF. Observa-se que, tanto em 2004 quanto em 2019, há grandes diferenças entre as unidades, variando, por exemplo, de apenas 2,3% de casos ausentes no Acre até 46,7% no Amapá, no primeiro ano. Em 2019, o percentual de ausências é maior, mas a amplitude entre as unidades é menor, indo de 12,4% no Paraná a 45,5% no Amapá.

TABELA 4

Estatísticas descritivas de declaração racial dos funcionários públicos por Unidade da Federação após a primeira estratégia de preenchimento de dados faltantes (2004 e 2019)

(Em %)

UF	2004			2019		
	Declarado	Não declarado	Total	Declarado	Não declarado	Total
Rondônia	94,5	5,5	100,0	82,1	17,9	100,0
Acre	97,8	2,3	100,0	73,6	26,4	100,0
Amazonas	95,2	4,8	100,0	68,7	31,3	100,0
Roraima	82,8	17,2	100,0	56,2	43,9	100,0
Pará	89,8	10,2	100,0	66,2	33,8	100,0
Amapá	53,3	46,7	100,0	54,5	45,5	100,0
Tocantins	92,7	7,4	100,0	74,1	25,9	100,0
Maranhão	89,4	10,6	100,0	66,2	33,8	100,0
Piauí	90,9	9,1	100,0	70,9	29,1	100,0
Ceará	85,8	14,2	100,0	74,8	25,2	100,0
Rio Grande do Norte	66,9	33,1	100,0	66,6	33,4	100,0
Paraíba	90,4	9,6	100,0	65,7	34,4	100,0
Pernambuco	94,4	5,7	100,0	78,4	21,6	100,0
Alagoas	71,2	28,8	100,0	67,3	32,7	100,0
Sergipe	89,5	10,5	100,0	79,9	20,1	100,0
Bahia	90,4	9,6	100,0	69,9	30,1	100,0
Minas Gerais	71,7	28,3	100,0	75,8	24,3	100,0
Espírito Santo	64,9	35,1	100,0	70,0	30,1	100,0
Rio de Janeiro	67,8	32,3	100,0	71,7	28,3	100,0
São Paulo	76,7	23,3	100,0	81,2	18,8	100,0
Paraná	96,2	3,8	100,0	87,6	12,4	100,0
Santa Catarina	92,9	7,1	100,0	86,4	13,6	100,0
Rio Grande do Sul	70,8	29,3	100,0	76,6	23,4	100,0
Mato Grosso do Sul	96,5	3,5	100,0	85,9	14,1	100,0
Mato Grosso	95,0	5,0	100,0	80,9	19,1	100,0
Goiás	69,8	30,2	100,0	72,5	27,5	100,0
Distrito Federal	65,0	35,1	100,0	69,3	30,7	100,0
Total	79,8	20,2	100,0	75,4	24,6	100,0

Fontes: Atlas do Estado Brasileiro (2022); e Rais (2004; 2019).

As UFs têm um papel importante para a imputação da raça na Rais por dois motivos. Em primeiro lugar, como se observa na tabela 4, a incidência de casos ausentes varia por unidade, sendo um importante identificador de padrões das ausências. Em segundo lugar, a UF será uma importante preditora, uma vez que a probabilidade de imputação de uma ou outra categoria racial ocorre em função da composição racial local. Ou seja, em estados com maior percentual

de população parda, por exemplo, é mais provável que o caso imputado seja pardo. Desse modo, trata-se de uma variável importante no processo de imputação.¹³

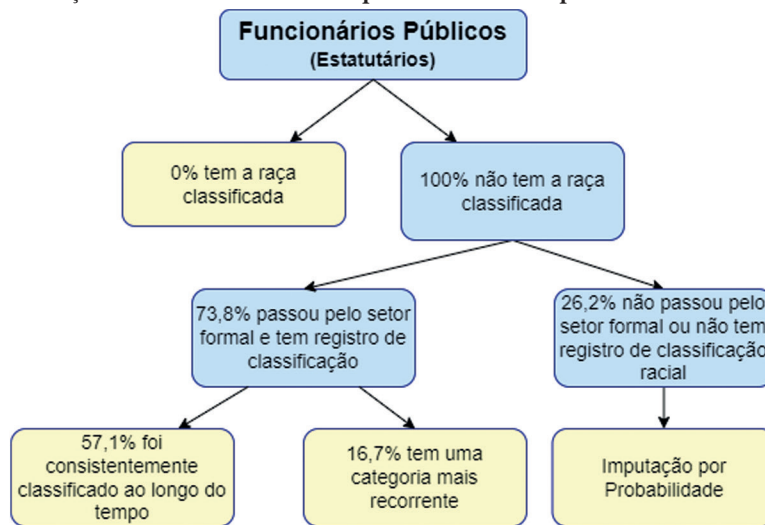
4.3 Síntese da estratégia de classificação e imputação da variável racial

A estratégia para contornar o alto percentual de não declaração racial no setor público brasileiro é sintetizada de acordo com o fluxograma da figura 1. Como se observa, há quatro caminhos para se chegar à classificação final a ser utilizada. O caminho *a* é daqueles trabalhadores que tiveram sua raça classificada no ano de referência – nesse caso, 2019. No caso dos estatutários, nenhum servidor foi classificado.

Entre aqueles que não possuem classificação, três caminhos são possíveis. O caminho *b* é daqueles que não possuem classificação em 2019, mas passaram em algum momento pelo setor privado e tiveram sua raça classificada consistentemente, isto é, foram classificados na mesma categoria em todos os anos. A esses é aferida a categoria em que foram classificados, correspondendo a 57,1% dos estatutários em 2019.

FIGURA 1

Fluxograma da situação classificatória no setor público brasileiro para a Rais



Elaboração do autor.

Obs.: Figura cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

O terceiro caminho, denominado *c*, é de servidores que já foram empregados no setor privado e tiveram a raça classificada em mais de uma categoria. Para esses, adota-se a classificação mais recorrente ao longo do tempo, e eles representam 16,7%. Por último, o caminho *d* representa 26,2% dos servidores para os quais não há registro de classificação ao longo do tempo na Rais, e a sua classificação foi imputada por probabilidade, seguindo o Mice.

5 COMPOSIÇÃO E DIFERENCIAL DE RENDA RACIAL

Esta seção tem como objetivo descrever os resultados das estratégias de uso da variável raça na Rais para o setor público. Para isso, a análise é feita de três maneiras, conforme resumido a seguir.

- 1) Pela descrição da composição racial e dos diferenciais de renda para os grupos raciais com e sem imputação.
- 2) Pela comparação entre a Rais imputada com a nova variável de raça e o Siape, com os mesmos indivíduos identificados, em que se pode comparar a categoria atribuída em cada uma das bases.
- 3) Pela comparação entre essa base e a PNAD Contínua, com estatísticas agregadas para o setor formal.

Todas as análises são realizadas com base no ano de 2019 e apresentam diferentes contrapontos à estratégia utilizada para a Rais.

13. Em anexo são apresentados dois modelos logísticos para estimar a probabilidade de ausência de classificação racial para os anos de 2004 e 2019 segundo as covariáveis apresentadas nas tabelas 3 e 4.

5.1 Descrição da Rais: antes e depois da estratégia

A tabela 5 resume o processo de preenchimento da variável raça sem nenhuma estratégia, com a primeira estratégia de utilização de registros anteriores, e, por fim, a estratégia de imputação dos dados via probabilidade. Nesse caso, trata-se de todos os servidores estatutários ativos em todos os poderes e esferas no Brasil. Nos dados originais, nenhum servidor estatutário possui classificação racial.

Já com o uso da primeira estratégia, observa-se que 73,8% possuem algum registro. Desses, considerando apenas os casos válidos, quase 70% é de pessoas consideradas brancas, 28,1% pardas e 3,1% pretas. Após a segunda estratégia, essa proporção permanece equivalente por grupos raciais. O percentual de brancos cai 1,2%, enquanto o percentual de pardos aumenta 1,0%. Ou seja, desse modo, a segunda estratégia não altera a proporção dos grupos raciais se comparada à estratégia que considera o histórico de registro de cada servidor, decorrente da metodologia adotada pelas duas estratégias.

TABELA 5

Composição racial do setor público (estatutários) de acordo com a estratégia de imputação da variável raça/cor – Brasil (2019)

Raça/cor	Original (%)	Primeira estratégia (%)		Segunda estratégia (%)
	Total	Total	Apenas com declaração	Total
Indígena	0,0	0,5	0,6	0,6
Branca	0,0	49,6	67,2	66,0
Preta	0,0	2,3	3,1	3,2
Parda	0,0	20,7	28,1	29,1
Amarela	0,0	0,8	1,0	1,0
Não identificado	100,0	26,2	-	0,0
Total de servidores		6.404.078		

Fontes: Atlas do Estado Brasileiro (2022); e Rais (2019).

Na tabela 6, tem-se o salário médio e a razão salarial entre os grupos em relação ao grupo branco, dependendo da estratégia. De maneira geral, os salários médios aumentam da primeira para a segunda estratégia, evidenciando que aqueles sem nenhum registro anterior de raça/cor têm salários mais altos, com exceção do grupo amarelo. Contudo, embora haja uma diferença no salário, as razões salariais dos maiores grupos (branco, pardo e preto) permanecem estáveis com ambas as estratégias. Entre os grupos menores, os indígenas passam de uma vantagem de 3,7% para 6,2% em relação aos brancos. No caso dos amarelos, a vantagem salarial diminui de 117,8% para 111,1%.

TABELA 6

Salário médio e razão salarial por estratégia de imputação no setor público (estatutários) – Brasil (2019)

Raça/cor	Primeira estratégia		Segunda estratégia	
	Salário médio (R\$)	Razão salarial (%)	Salário médio (R\$)	Razão salarial (%)
Indígena	5.303,30	103,7	5.705,79	106,2
Branca	5.114,20	100,0	5.373,13	100,0
Preta	3.838,04	75,0	4.237,65	78,9
Parda	4.494,57	87,9	4.598,43	85,6
Amarela	6.023,38	117,8	5.969,90	111,1
Não declarado	5.735,88	112,2	-	-

Fontes: Atlas do Estado Brasileiro (2022); e Rais (2019).

Dessa maneira, os resultados indicam que, para o grupo dos servidores estatutários, a composição racial com a primeira ou a segunda estratégia é intercambiável. Isto é, ambas as estratégias resultam em uma divisão proporcional equivalente entre os grupos – o que seria de se esperar, já que a segunda estratégia é subsidiária da primeira.

No caso do diferencial de renda entre os grupos raciais, ambas as estratégias continuam sendo válidas para estimar a desigualdade. A imputação feita na segunda parte não altera substancialmente as razões salariais. Para validar a estratégia, contudo, é necessário contrastar os seus resultados com outras bases. Na subseção seguinte, são comparados

os resultados da Rais com o Siape, lembrando que se trata de formatos classificatórios semelhantes. Na outra subseção, a Rais é comparada com a PNAD Contínua, tratando-se, portanto, de duas formas distintas de classificação.

5.2 Rais e Siape

O Siape é outra fonte de dados com cobertura representativa do setor público – embora ela seja restrita à esfera federal. Como apontado e analisado por Silva e Lopez (2021), o percentual de não preenchimento da variável raça nessa base tem diminuído desde o final dos anos 1990, quando era de quase 20%. Já em 2019, os casos não preenchidos representavam apenas 5% do total de servidores. Trata-se de um registro administrativo que, em tese, é autodeclarado, uma vez que os servidores podem acessar o seu cadastro e alterá-lo de acordo com sua determinação racial. Contudo, não é tão bem definido como é feita a primeira inserção do dado racial no registro.

Uma vez que se tem disponível a identificação dos servidores no Siape e na Rais, é possível, portanto, cruzar a raça inserida no primeiro banco com a raça imputada no segundo. Trata-se de uma validação extremamente acurada da imputação realizada, uma vez que são comparados caso a caso. A sua análise, entretanto, deve ser feita com cuidados, levando em consideração que na maioria de *surveys* e pesquisas domiciliares em que há duas variáveis raciais coletadas (autoidentificada e heterodeclarada, por exemplo), a sobreposição raramente é completa (Silva, 1999a; Muniz, 2012; Silveira, 2019; Silveira e Tomás, 2019), cabendo compreender os motivos que levaram a tais diferenças.

Em todas essas comparações da subseção 5.2, é importante ressaltar que se trata sempre dos mesmos casos, e em alguns deles a raça é classificada em uma categoria no Siape, mas imputada em outra na Rais. Esta subseção, portanto, se dedica a analisar as consistências e a compreender as inconsistências da técnica de imputação, ressaltando, novamente, que se trata do universo de casos da esfera federal registrados em 2019.

Na tabela 7, tem-se a distribuição das categorias raciais do Siape, com e sem os casos não declarados, e as categorias da Rais. Em primeiro lugar, pode-se identificar que 5,2% dos casos no Siape não foram classificados em 2019, como já apontado por Silva e Lopez (2021). Já na segunda e na terceira coluna, distribuem-se os casos válidos das duas bases, em que se atesta o perfil “mais branco” da Rais. Nesta, 64,5% dos casos são considerados brancos em comparação a 58,7% no Siape. Todas as demais categorias são ligeiramente menores na Rais, sendo o grupo pardo aquele com maior diferença, em 2,9%.

TABELA 7

Distribuição racial da esfera federal para o Siape e a Rais com categorias imputadas – Brasil (2019)

(Em %)

Categorias	Siape	Siape (válidos)	Rais imputada
Amarela	2,4	2,5	1,8
Branca	55,6	58,7	64,5
Indígena	0,3	0,3	0,6
Parda	30,9	32,6	29,7
Preta	5,6	5,9	3,4
Não informada	5,2	-	-
Total	100,0	100,0	100,0

Fontes: Atlas do Estado Brasileiro (2022); Rais (2019); e Siape (2019).

Na tabela 8, observa-se que a consistência da imputação na Rais, se comparada ao Siape, varia consideravelmente por categoria racial. De maneira geral, 69,9% dos casos (quando desconsiderados os casos não informados no Siape) são consistentemente classificados no Siape e na Rais com categorias imputadas. Essa precisão, contudo, é muito baixa no caso das categorias amarela, indígenas e preta. Os casos tidos como amarelos no Siape são assim classificados em apenas 30,8% na Rais, e a maioria dos casos é predita como branca e 21,4%, como parda. Situação semelhante é observada no caso dos indígenas, em que apenas 28,9% são consistentemente preditos, se considerarmos o Siape como baliza. No caso da categoria preta, os casos se distribuem entre um terço como brancos, um terço como pardos e o outro terço como pretos.

TABELA 8

Cruzamento das categorias raciais da esfera federal do Siape e da Rais imputada – Brasil (2019)

(Em %)

Siape	Raça imputada (Rais)					
	Amarela	Branca	Indígena	Parda	Preta	Total
Amarela	30,8	45,7	0,7	21,4	1,5	100,0
Branca	0,9	82,9	0,4	14,9	0,9	100,0
Indígena	1,9	39,6	28,9	27,5	2,1	100,0
Parda	1,5	38,9	0,6	56,4	2,6	100,0
Preta	1,0	31,4	0,5	33,6	33,5	100,0
Não informado	1,2	64,5	0,8	30,0	3,6	100,0
Total	1,8	64,5	0,6	29,7	3,4	100,0

Fontes: Atlas do Estado Brasileiro (2022); Rais (2019); e Siape (2019).

Em caso de restrição das categorias aos grupos preto, pardo e branco, observa-se um aumento na consistência das duas bases. Ou seja, excluindo-se as categorias amarela e indígena, que tiveram menor consistência pela predição, e agregando as categorias preta e parda como negra, a consistência aumenta de 69,9% para 75,1% dos casos. Como se observa na tabela 9, 61,5% dos casos considerados como negros no Siape também são assim considerados na imputação da Rais, enquanto para os brancos esse percentual sobe para 84,0%.

TABELA 9

Cruzamento das categorias raciais da esfera federal do Siape e da Rais imputada agregadas em negra e branca – Brasil (2019)

(Em %)

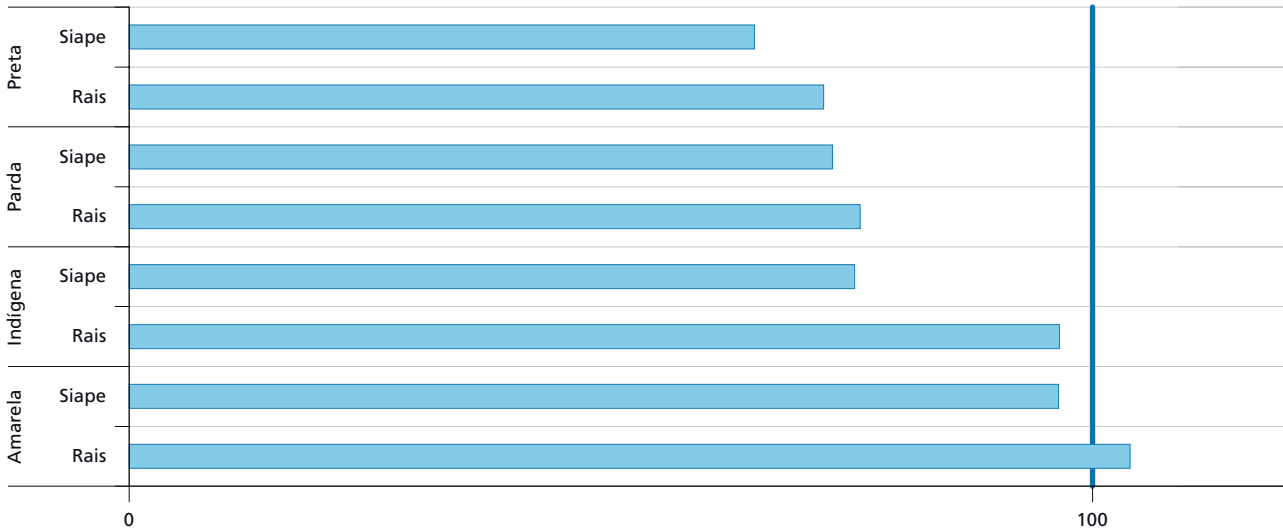
Siape	Raça imputada (Rais)		
	Negra	Branca	Total
Negra	61,5	38,5	100,0
Branca	16,0	84,0	100,0
Total	33,9	66,1	100,0

Fontes: Atlas do Estado Brasileiro (2022); Rais (2019); e Siape (2019).

Por fim, os gráficos 1 e 2 são dedicadas a analisar as diferenças salariais por categoria racial. Utilizou-se, para essa finalidade, a distribuição racial do Siape e da Rais imputada, uma vez que a variável de rendimentos é a mesma para ambos os casos – proveniente da Rais. Além disso, tem-se a razão salarial em percentual, tomando o grupo branco como referência. Na tabela 8, tem-se uma diferença considerável dos valores dos rendimentos para as categorias amarela, indígena e preta. A diferença entre os valores médios, considerando a Rais em vez do Siape, diminui o salário desses grupos em R\$ 443,00, R\$ 1.996,00 e R\$ 534,00, respectivamente. Já a razão salarial tem uma diferença de 7,4 pontos percentuais (p.p.) para amarelos, 21,3 p.p. para indígenas e 7,2 p.p. para pretos. Por seu turno, para brancos o valor da média se altera em R\$ 355,00, e apenas R\$ 50,00 para pardos.

GRÁFICO 1

Razão salarial para as categorias raciais do Siape e da Rais imputada na esfera federal – Brasil (2019)

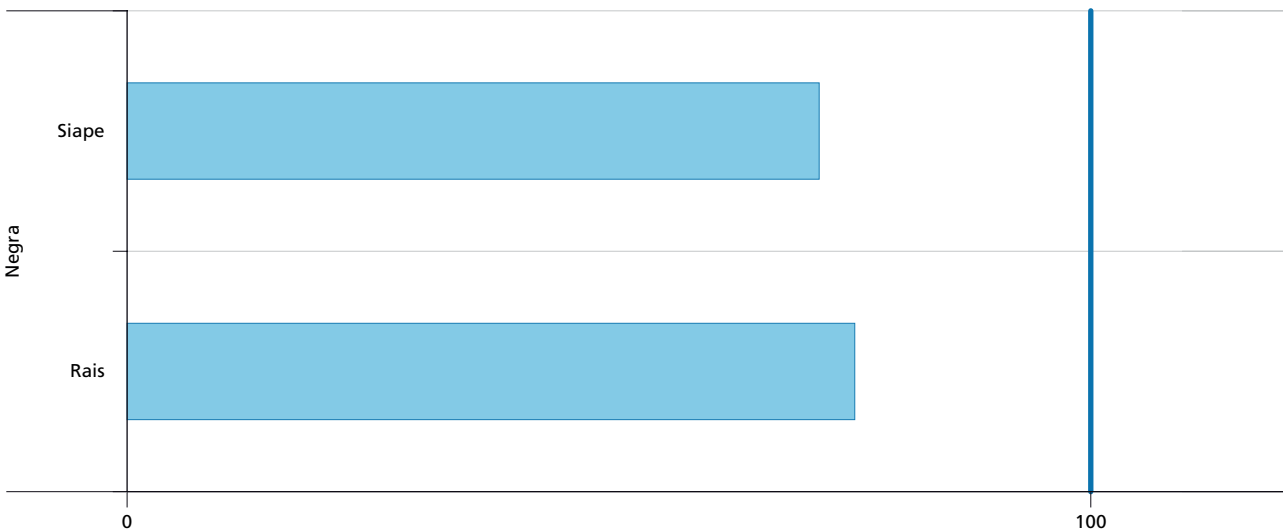


Fontes: Atlas do Estado Brasileiro (2022); Rais (2019); e Siape (2019).

Considerando que a estratégia de agregar as categorias preta e parda em negra levou a resultados mais consistentes da imputação, a tabela 9 apresenta o resultado das diferenças salariais apenas para este recorte. Neste caso, observa-se que a diferença entre a média salarial da Rais imputada é de R\$ 144,00 para os servidores negros, e a razão salarial é apenas 3,8 p.p. menor, o que torna o resultado dessa comparação menos dispar.

GRÁFICO 2

Razão salarial para as categorias raciais do Siape e da Rais imputada agregadas em negra e branca na esfera federal – Brasil (2019)



Fontes: Atlas do Estado Brasileiro (2022); Rais (2019); e Siape (2019).

Desse modo, o uso do Siape como referência para analisar os resultados da imputação das categorias raciais revela limitações das estratégias. Como se observa, as categorias branca e parda são maiores e as imputações são mais recorrentes nessas categorias, de maneira que grande parte dos casos considerados amarelos, pretos ou indígenas são imputados nessas duas categorias. Em especial, as categorias amarela e indígena levam a distribuições e a médias salariais muito díspares, se considerarmos como referência o Siape. No caso da categoria preta, a diferença é menor, e ainda conta com a possibilidade de agregá-los aos pardos na categoria negra, reduzindo as distorções.

5.3 Comparação entre Rais e PNAD Contínua

A PNAD Contínua é a fonte mais usual de estatísticas populacionais no Brasil, devido à sua representatividade e à sua periodicidade trimestral. Em relação à variável raça, a PNAD Contínua conta com um procedimento de autodeclaração do entrevistado, embora esse declare a classificação racial dos outros membros do domicílio. Desse modo, utilizar essa

fonte de dados para comparação se mostra válido tanto pela sua difusão quanto por sua forma de declaração ser um contraponto àquela realizada na Rais.

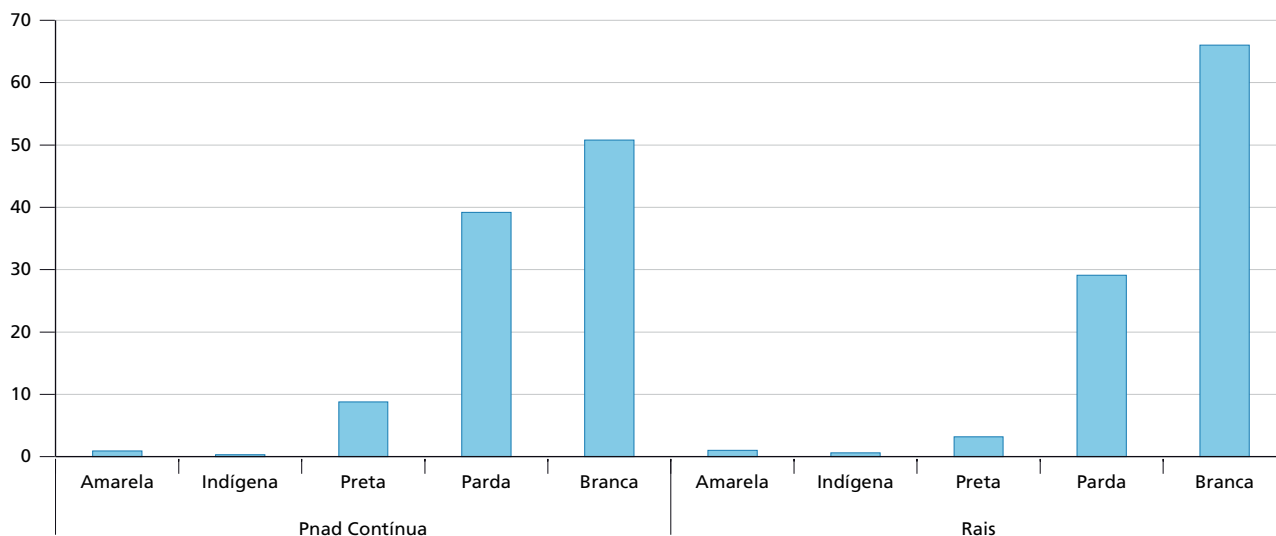
Paixão *et al.* (2012) utilizam a PNAD (no formato anual) para apontar limites e possibilidades de comparação com a Rais, em termos de classificação Rais. Neste trabalho, o setor público recebe maior atenção, além do fato de se tratar de uma validação da estratégia de preenchimento dos casos não preenchidos.

Para isso, restringiu-se a amostra da PNAD Contínua aos trabalhadores estatutários de todos os poderes e esferas no ano de 2019, assim como a amostra retirada da Rais. Os resultados no gráfico 3 demonstram, como também feito por Paixão *et al.* (2012), uma sobre-representação de trabalhadores brancos na Rais.

GRÁFICO 3

Composição racial do setor público (estatutários) na Rais e PNAD Contínua – Brasil (2019)

(Em %)

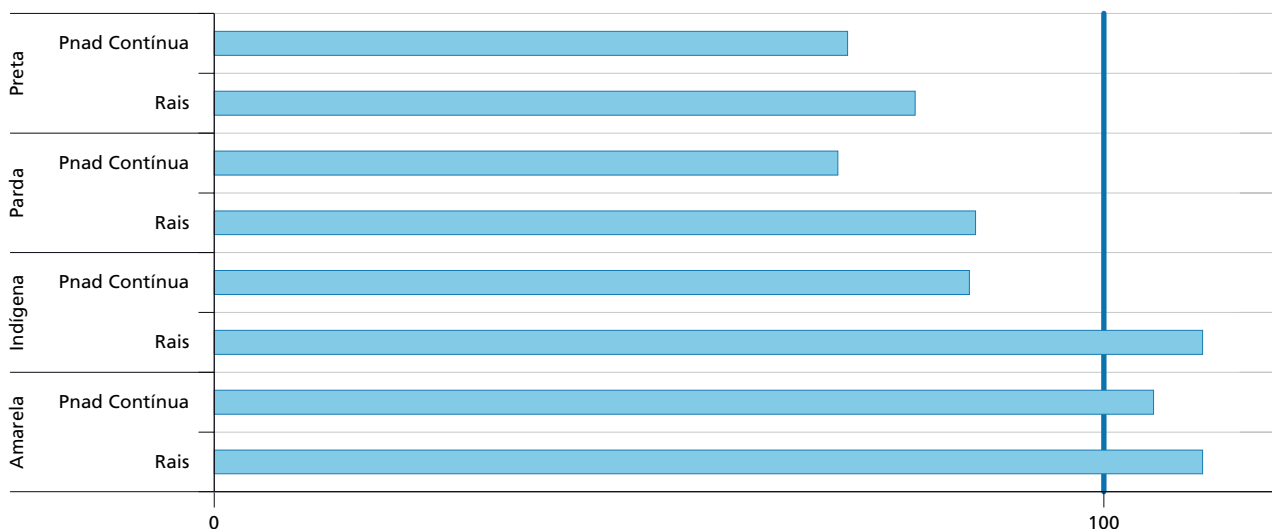


Fontes: Atlas do Estado Brasileiro (2022); Rais (2019); e PNAD Contínua.

Conforme o gráfico 3 indica, a Rais imputada possui 15,2% a mais de servidores classificados como brancos do que a PNAD Contínua. A composição de trabalhadores pretos na Rais é menor que a metade estimada pela PNAD Contínua, perfazendo apenas 3,2% dos casos. Já a população parda, que é de 39,2% na amostra domiciliar, cai para 29,1% nos registros administrativos heterodeclarados. Os indígenas, por sua vez, representam 0,3% e 0,6%; e os amarelos, 0,9% e 1,0% na PNAD Contínua e na Rais, respectivamente.

Contudo, também como analisado por Paixão *et al.* (2012), a distorção da estimativa da desigualdade é menor do que a composição racial. No gráfico 4, observa-se que na PNAD Contínua, em média, uma pessoa preta estatutária recebe 71,2% do salário de uma pessoa branca na mesma situação contratual, enquanto na Rais essa razão percentual é de 78,8%. No caso do grupo pardo, a diferença é maior. De acordo com a PNAD Contínua, o salário médio desse grupo é 70,1% do grupo branco, enquanto na Rais essa razão é de 85,6%, em 2019.

O grupo indígena é o que apresenta maiores variações, e merece maior atenção. Considerando a PNAD Contínua, uma pessoa indígena em regime estatutário recebe 84,9% do que recebe uma pessoa branca. Já na Rais, essa razão sobe para 111,1%, ou seja, em média, esse grupo recebe mais que o grupo de referência. Nesse caso, não se trata apenas de uma diferença de intensidade, mas também de sentido da desigualdade, pelo menos de acordo com os valores médios. O grupo amarelo, por sua vez, apresenta pouca variação, sendo positivo em ambos – 105,6% na PNAD Contínua e 111,1% na Rais.



Fontes: Atlas do Estado Brasileiro (2022); Rais (2019); e PNAD Contínua (2019).

Com exceção da estimativa da razão salarial entre grupos indígenas/brancos, a comparação entre Rais e PNAD Contínua em 2019 apresenta pontos de convergência. Como sustentado por Paixão *et al.* (2012), mesmo que haja subestimação de diferenciais, ela não inverte o resultado quando se trata de pretos, pardos e brancos – e como se observou no gráfico 4, também para amarelos. Sendo assim, a despeito das diferenças de forma de classificação da variável racial – auto e heteroclassificação – e pelo fato de se tratar de uma amostra e de um universo, o uso da PNAD Contínua e da Rais não causa viés nas estimativas, e são comparáveis em termos de análise de desigualdades.

6 CONSIDERAÇÕES E RECOMENDAÇÕES DE USO

O uso da variável raça na Rais exige uma série de cuidados (Paixão *et al.*, 2012; Noronha, 2018; Silveira, 2019; Dias e Rocha, 2021; Miller e Schmutte, 2021). O pesquisador deve ter em mente que se trata de uma representação da dimensão racial, mas que o fato de o formato ser diferente daqueles usados pelo IBGE leva a distorções de composição e estimativas de desigualdade. Como realizado por Paixão *et al.* (2012), é necessário pontuar os limites desse uso. Ao mesmo tempo, a variável raça está presente na base, e possui vários pontos de toque com as formas autodeclaradas de mensuração. Portanto, esta nota teve como objetivo: i) apresentar uma proposta de estratégia metodológica para contornar o não preenchimento da variável; e ii) demonstrar até que ponto elas convergem com outras fontes de dados usuais nas análises de desigualdade e políticas públicas no Brasil. Ter a variável raça/cor disponível na Rais para o setor público permitiria um ganho em termos de representatividade estatística para análises desagregadas por poderes, esferas e entes subnacionais.

Primeiramente, em relação às estratégias metodológicas para contornar as não declarações raciais, pontua-se que a maior parcela de redução dos dados faltantes ocorre por meio do resgate de registros anteriores. Como demonstrado, o percentual de *missing* diminui de 100,0% para 26,2%, apenas com a primeira parte da estratégia. Essa estratégia já foi utilizada por Noronha (2018) e Silveira (2019). Nesse sentido, os pontos de atenção com relação à essa estratégia são relacionados à heteroclassificação e ao fato de esse ato ser realizado por agentes difusos, e não por entrevistadores treinados, como é o caso de *surveys* e pesquisas amostrais.

Após essa primeira etapa, a imputação por meio do Mice é realizada, e preenche os 26,2% de casos que não foram preenchidos com a etapa anterior. Nessa, a imputação é feita por probabilidade, de acordo com as características do indivíduo e do local em que ele está inserido. Logo, o resultado final é associado às características do universo de análise aos quais aqueles casos estão submetidos. Por exemplo, se a Rais tende a sobrerrepresentar a classificação branca, a imputação também será permeada por essa característica geral das classificações. Em geral, os resultados são probabilísticos e não enviesam os resultados – isto é, o viés existente “contamina” a imputação, mas não cria um novo viés.

Desse modo, o uso da variável raça conforme o formato proposto deve levar isso em consideração. As comparações realizadas da Rais imputada com Siape e com a PNAD Contínua demonstram até que ponto se pode assumir

sobreposição das informações entre elas para análise de dados do mercado de trabalho e, especialmente, sobre o funcionalismo público brasileiro.

Em relação ao Siape, o uso da variável imputada de raça na Rais apresenta vários pontos de atenção. No formato adotado, a proporção de pessoas pretas na esfera federal se tornou menor, caindo de 5,3% no Siape para 3,3% na Rais. Ou seja, se não há diferenças na estimativa de desigualdade salarial, as estimativas de inserção de acesso e inserção de servidores pretos podem ser subestimadas.

Para as categorias amarela, indígena e preta, as sobreposições são inferiores a 50%, o que indica uma baixa correspondência entre as duas bases. Nesse sentido, a estratégia adotada não é suficiente e confiável para essas categorias, embora sejam para as categorias *parda* e *branca*. Contudo, ao excluir as categorias *amarela* e *indígena* e agregar pretos e pardos na categoria *negra*, o percentual de sobreposição aumenta consideravelmente, e as estimativas de diferencial salarial se tornam equivalentes entre Siape e Rais.

As diferenças também ficam elucidadas na comparação com a PNAD Contínua. Em relação a esta, tem-se como pontos positivos que as estimativas de desigualdade salarial indicam os mesmos sentidos, novamente, com exceção do grupo indígena. Para os grupos preto e pardo, a Rais subestima a vantagem do grupo branco – em 7,7% e 15,4%, respectivamente –, enquanto superestima a vantagem dos amarelos – em 5,5%.

Portanto, a imputação da variável racial na Rais a torna viável para análises temporais do setor público. Porém, isso não é feito sem ressalvas, e pesquisadores e gestores que utilizem esses dados devem ter em mente o limite das estratégias. O uso da imputação neste estágio, exige alguns cuidados, como os descritos a seguir.

- 1) Excluir os casos imputados para amarelos e indígenas, uma vez que os modelos não levam a um alto percentual de consistência dessas categorias.
- 2) A agregação das categorias *preta* e *parda* em *negra* pode levar a uma maior consistência das imputações e a estimativas mais confiáveis de diferencial salarial.
- 3) A ênfase em análises agregadas é mais confiável do que a ênfase naquelas em que se analisam aspectos de trajetória individual.
- 4) E no caso de modelos multivariados, o controle por meio de uma variável auxiliar que indique que se trata de casos imputados aumenta o rigor para a análise dos resultados provenientes desses casos.

As estratégias de contornar a ausência de declaração racial na Rais devem ser compreendidas como um caminho ainda a ser aprimorado. Como sugestão, aponta-se que esses dados estão disponíveis para anos anteriores a 2019, e estimulam-se novas explorações e análises, incluindo aspectos comparativos com outras bases e estudos já realizados. Outros recortes podem ser feitos, como é o caso dos trabalhadores que não estão sob regime estatutário – esses foram utilizados apenas para se fazerem comparações mais homogêneas.

REFERÊNCIAS

- ALBIERI, S. **A ausência de resposta em pesquisas**: uma aplicação de métodos de imputação. 1989. 138 f. Dissertação (Mestrado) – Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1989.
- ALMEIDA, L. A. D. **Deficiência e desigualdades no Brasil**: pobreza, inserção no mercado de trabalho e renda. 2019. 381 f. Tese (Doutorado) – Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais, 2019.
- AZUR, M. J. *et al.* Multiple imputation by chained equations: what is it and how does it work? **International Journal of Methods in Psychiatric Research**, v. 20, n. 1, p. 40-49, Mar. 2011.
- BAILEY, S.; LOVEMAN, M.; MUNIZ, J. Measures of “race” and the analysis of racial inequality in Brazil. **Social Sciences Research**, v. 42, n. 1, p. 106-119, Jan. 2013.
- BRAZ, R. *et al.* Avaliação da completude da variável raça/cor nos sistemas nacionais de informação em saúde para aferição da equidade étnico-racial em indicadores usados pelo Índice de Desempenho do Sistema Único de Saúde. **Saúde em Debate**, Rio de Janeiro, v. 37, n. 99, p. 554-562, out.-dez. 2013.
- CARVALHO, J.; WOOD, C.; ANDRADE, F. Estimating the stability of census-based racial/ethnic classifications: the case of Brazil. **Population Studies**, London, v. 58, n. 3, p. 331-343, Nov. 2004.
- CORNWELL, C.; RIVERA, J.; SCHMUTTE, I. Wage discrimination when identity is subjective. **Journal of Human Resources**, v. 52, n. 3, p. 719-755, 2017.

- DE MICHELI, D. A. **Back to black**: racial reclassification and political identity formation in Brazil. 2019. 305 f. Dissertation (Doctoral Degree) – Faculty of the Graduate School, Cornell University, Ithaca, 2019.
- DIAS, M.; ROCHA, R. **Racial gaps in the labor market**: the role of nonwhite entrepreneurship. Rochester: SSRN, 2021. Disponível em: <<https://bit.ly/3Tc2WsW>>.
- HEGDE, H. *et al.* Mice vs PPCA: missing data imputation in healthcare. **Informatics in Medicine Unlocked**, v. 17, p. 100275, 2019.
- IGNATIEV, N. (Ed.). **How the Irish became white**. New York: Routledge, 1995.
- LOPEZ, F.; GUEDES, E. **Três décadas de evolução do funcionalismo público no Brasil (1986-2017)**. Brasília: Ipea, ago. 2020. (Texto para Discussão, n. 2579).
- LOVEMAN, M.; MUNIZ, J. How Puerto Rico became white: boundary dynamics and intercensus racial reclassification. **American Sociological Review**, v. 72, n. 6, p. 915-939, Dec. 2007.
- MILLER, C.; SCHMUTTE, I. M. **The dynamics of referral hiring and racial inequality**: evidence from Brazil. Cambridge: NBER, Sept. 2021. (Working Paper, n. 29246).
- MIRANDA, V. A resurgence of black identity in Brazil? Evidence from analysis of recent censuses. **Demographic Research**, v. 32, n. 59, p. 1603-1630, Jun. 2015.
- MONK, E. **Color, bodily capital, and ethnoracial division in the U.S. and Brazil**. 2013. 176 f. Dissertation (Doctoral Degree) – University of California, Berkeley, 2013.
- MUNIZ, J. Preto no branco? Mensuração, relevância e concordância classificatória no país da incerteza racial. **Dados**, Rio de Janeiro, v. 55, n. 1, p. 251-282, 2012.
- NAKAMURA, L. M.; VAZ, D. V. Quanto heterogêneo é o setor público no Brasil? Uma análise das diferenças salariais entre os poderes Executivo, Legislativo e Judiciário. **Administração Pública e Gestão Social**, v. 12, n. 4, p. 1-24, out.-dez. 2020.
- NORONHA, C. L. A.; VILELA, E. M. Análise comparativa da mobilidade ocupacional entre homens e mulheres estrangeiros no mercado formal de trabalho brasileiro. **Revista Brasileira de Sociologia**, v. 8, n. 19, p. 148-177, maio-ago. 2020.
- NUNES, W.; TELES, J. A elite salarial do funcionalismo público federal: sugestões para uma reforma administrativa mais eficiente. **Cadernos de Gestão Pública e Cidadania**, v. 26, n. 84, p. 1-24, maio 2021.
- OSÓRIO, R. **O sistema classificatório de “cor ou raça” do IBGE**. Brasília: Ipea, nov. 2003. (Textos para Discussão, n. 996). Disponível em: <<https://bit.ly/3yLA7v7>>.
- PAIXÃO, M. *et al.* Investigação sobre qualidade da variável cor ou raça na Rais através de um estudo comparativo com a PNAD do IBGE. In: ENCONTRO ANUAL DA ANPOCS, 36., 2012, Águas de Lindoia, São Paulo. **Anais... Águas de Lindoia**: Anpocs, 2012.
- PETRUCCELLI, J. L. Autoidentificação, identificação étnico-racial e heteroclassificação. In: PETRUCCELLI, J. L.; SABOIA, A. L. (Org.). **Características étnico-raciais da população**: classificações e identidades. Rio de Janeiro: IBGE, 2013. p. 31-50. Disponível em: <<https://bit.ly/3D45iV6>>.
- RIBEIRO, C. A. C. Contínuo racial, mobilidade social e embranquecimento. **Revista Brasileira de Ciências Sociais**, v. 32, n. 95, p. 1-25, 2017.
- ROTH, W. D. Racial mismatch: the divergence between form and function in data for monitoring racial discrimination of Hispanics. **Social Science Quarterly**, v. 91, n. 5, p. 1288-1311, Dec. 2010.
- SANSON, J. R.; MOUTINHO, L. M. de G. A evolução do emprego público no Brasil: 1950-80. **Ensaios FEE**, v. 8, n. 2, p. 31-48, 1987.
- SANTOS, C. H. M. dos *et al.* O que pode dar errado quando se usam os microdados da Rais para análises do emprego público por ente federado? Um inventário dos problemas conhecidos e algumas soluções. **Mercado de Trabalho: conjuntura e análise**, Brasília, n. 65, p. 39-56, out. 2018.
- SAPERSTEIN, A. Capturing complexity in the United States: which aspects of race matter and when? **Ethnic and Racial Studies**, v. 35, n. 8, p. 1484-1502, 2012.
- SAPERSTEIN, A.; PENNER, A. Racial fluidity and inequality in the United States. **American Journal of Sociology**, v. 118, n. 3, p. 676-727, Nov. 2012.
- SENKEVICS, A. De brancos para negros? Uma análise longitudinal da reclassificação racial no Enem 2010-2016. **Dados**, Rio de Janeiro, v. 65, n. 3, p. 1-40, 2022. Disponível em: <<https://bit.ly/3ezPxM2>>.

- SENKEVICS, A.; MACHADO, T.; OLIVEIRA, S. **A cor ou raça nas estatísticas educacionais**: uma análise dos instrumentos de pesquisa do Inep. Brasília: Inep, 2016. (Textos para Discussão, n. 41). Disponível em: <<https://bit.ly/3VBjAU8>>.
- SILVA, N. V. Nota sobre “raça social” no Brasil. *In*: HASENBALG, C.; SILVA, N. V.; LIMA, M. (Org.). **Cor e estratificação social**. Rio de Janeiro: Contra Capa, 1999a. p. 107-125.
- _____. Morenidade: modos de usar. *In*: HASENBALG, C.; SILVA, N. V.; LIMA, M. (Org.). **Cor e estratificação social**. Rio de Janeiro: Contra Capa, 1999b. p. 86-106.
- SILVA, T. D.; LOPEZ, F. G. **Perfil racial do serviço civil ativo do Executivo federal (1999-2020)**. Brasília: Ipea, jul. 2021. (Nota Técnica Diest, n. 49).
- SILVEIRA, L. S. **Reclassificação racial e desigualdade**: análise longitudinal de variações socioeconômicas e regionais no Brasil entre 2008 e 2015. 2019. 151 f. Tese (Doutorado) – Faculdade de Filosofia e Ciências Humanas, Universidade Federal de Minas Gerais, Belo Horizonte, 2019.
- SILVEIRA, L. S.; TOMAS, M. C. Fluidez racial na Região Metropolitana de Belo Horizonte: características individuais e contexto local na construção da raça. **Revista Brasileira de Estudos de População**, v. 36, p. 1-22, 2019.
- TELLES, E. E. Who are the morenas? **Social Forces**, v. 73, n. 4, p. 1609-1611, Jun. 1995.
- TELLES, E.; PASCHEL, T. Who is black, white and mixed race? How skin color, status and nation shape racial classification in Latin America. **American Journal of Sociology**, v. 120, p. 3, p. 864-907, Nov. 2014.
- VARGAS, N. Off white: colour-blind ideology at the margins of whiteness. **Ethnic and Racial Studies**, v. 37, n. 13, p. 2281-2302, 2014.
- ZHANG, G. *et al.* Multiple imputation of missing race and ethnicity in CDC covid-19 case-level surveillance data. **International Journal of Statistics in Medical Research**, v. 11, p. 1-11, Jan. 2022.

BIBLIOGRAFIA COMPLEMENTAR

- SILVEIRA, L. S.; PENHA, P. H. E. da; CASTRO, B. P. C. de. Funcionalismo público brasileiro no século XXI: transformações ocupacionais nos diferentes níveis do poder executivo entre 2003 e 2018. **Cadernos Gestão Pública e Cidadania**, São Paulo, v. 26, n. 84, p. 1-20, 2021.
- THOMPSON, C. A. *et al.* A multiple-imputation “forward bridging” approach to address changes in the classification of Asian race/ethnicity on the US death certificate. **American Journal of Epidemiology**, v. 187, n. 2, p. 347-357, Feb. 2018.

Ipea – Instituto de Pesquisa Econômica Aplicada

EDITORIAL

Coordenação

Aeromilson Trajano de Mesquita

Assistentes da Coordenação

Rafael Augusto Ferreira Cardoso

Samuel Elias de Souza

Supervisão

Camilla de Miranda Mariath Gomes

Everson da Silva Moura

Revisão

Alice Souza Lopes

Amanda Ramos Marques

Ana Clara Escórcio Xavier

Barbara de Castro

Clícia Silveira Rodrigues

Olavo Mesquita de Carvalho

Regina Marta de Aguiar

Reginaldo da Silva Domingos

Brena Rolim Peixoto da Silva (estagiária)

Nayane Santos Rodrigues (estagiária)

Editoração

Anderson Silva Reis

Cristiano Ferreira de Araújo

Danielle de Oliveira Ayres

Danilo Leite de Macedo Tavares

Leonardo Hideki Higa

Capa

Aline Cristine Torres da Silva Martins

Projeto Gráfico

Danielle de Oliveira Ayres

Flaviane Dias de Sant'ana

*The manuscripts in languages other than Portuguese
published herein have not been proofread.*

Ipea – Brasília

Setor de Edifícios Públicos Sul 702/902, Bloco C

Centro Empresarial Brasília 50, Torre B

CEP: 70390-025, Asa Sul, Brasília-DF

Missão do Ipea

Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.

ipea Instituto de Pesquisa
Econômica Aplicada

MINISTÉRIO DA
ECONOMIA



PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL