

**Título:** A utilização da RAIS em estudos longitudinais para análise da mobilidade ocupacional de estrangeiros no mercado de trabalho.

## **1. Resumo**

Essa pesquisa tem como objetivo discutir a utilização da Relação Anual de Informações de Sociais (RAIS), base de dados administrativa do Ministério do Trabalho e Emprego (MTE) em formato longitudinal de painel para compreensão da trajetória ocupacional dos estrangeiros empregados formalmente no Brasil. A RAIS é o único banco de dados brasileiro público que torna possível a análise longitudinal de painel sobre os indivíduos no mercado de trabalho formal, na medida em que dispõe de informações que possibilitam a identificação dos indivíduos como o número de inscrição no Programa de Integração Social (PIS) de cada trabalhador e do Cadastro Nacional de Pessoa Jurídica (CNPJ). Entretanto, a RAIS apresenta inúmeras dificuldades de utilização, visto que conta ainda com um número bastante restrito de usuários. As principais fragilidades desse banco são: 1) eventuais imprecisões nos dados da RAIS, decorrentes da utilização de “registros administrativos que, pela sua própria natureza, não sofreram qualquer processo de crítica; 2) os registros contidos na RAIS são restritos ao mercado formal, inviabilizando, dessa forma, a apreensão da situação de trabalhadores que não estejam formalmente empregados; 3) Em específico sobre os estrangeiros, a RAIS a variável “nacionalidade”, na qual se obtém o país de nascimento do estrangeiro é modificada ao longo dos anos. Nesse sentido, esse artigo tem como objetivo discutir estratégias metodológicas com intuito de contribuir para a superação dessas fragilidades existentes na RAIS e ampliar o conhecimento de mais pesquisadores sobre essa importante fonte de dados, permitindo que o aumento de usuários dessas informações.

## 2. Introdução

Essa pesquisa tem como objetivo discutir a utilização da Relação Anual de Informações de Sociais (RAIS), base de dados administrativa do Ministério do Trabalho e Emprego (MTE) em formato longitudinal de painel para compreensão da trajetória ocupacional dos estrangeiros empregados formalmente no Brasil.

A RAIS é a única base de dados brasileira pública que torna possível a análise longitudinal de painel sobre os indivíduos no mercado de trabalho formal, na medida em que dispõe de informações sobre o número de inscrição no Programa de Integração Social (PIS) de cada trabalhador e do Cadastro Nacional de Pessoa Jurídica (CNPJ) que identificam os indivíduos e as empresas a cada ano, respectivamente. Isto é, tratada isoladamente, a RAIS é também uma base de tipo *cross-section*, mas associando os diversos anos por meio do PIS do indivíduo, torna-se possível a construção de um censo dos trabalhadores inseridos no mercado de trabalho formal em formato longitudinal de painel. Dessa forma, essa base de dados propicia o acompanhamento, ao longo do tempo, da trajetória dos indivíduos, através de seus vínculos de trabalho no mercado formal, contemplando as dimensões geográfica, setorial e ocupacional.

Entretanto, a RAIS conta ainda com um número bastante restrito de usuários, tendo em vista seu potencial. As informações também apresentam inúmeras dificuldades de utilização, sobretudo em razão da falta de padronização do formato entre os anos. Ou seja, há eventuais imprecisões nos dados da RAIS, decorrentes da utilização de “registros administrativos que, pela sua própria natureza, não sofreram qualquer processo de crítica, sendo, por isso mesmo, passíveis de apresentarem erros provenientes de declarações mal preenchidas” (Sternberg, 2000, p. 186). O preenchimento dos dados é realizado pelo próprio empregador (ou pessoa responsável por tal atividade na empresa) que, por muitas vezes, possivelmente declara dados não atualizados e incorretos sobre o trabalhador.

Além dessa limitação inerente à origem dos dados, é preciso considerar também outra, particularmente importante para a análise de trajetórias ocupacionais, que diz respeito à cobertura dos dados, uma vez que os registros nela contidos são restritos ao mercado formal, inviabilizando, dessa forma, a apreensão da situação de trabalhadores que não estejam formalmente empregados. Assim, quando o trabalhador deixa de ser captado pela base, é impossível saber se ele permanece

ocupado — neste caso, no mercado informal —, se está desempregado ou na inatividade.

Em específico sobre os estrangeiros, a RAIS mostra-se uma importante base de dados a ser utilizada visto que apresenta a informação do país de origem do trabalhador, tornando possível a compreensão da dinâmica de trabalho dos estrangeiros no Brasil. Essa informação é obtida por meio da variável “nacionalidade”. Entretanto, uma dificuldade relacionada a essa informação refere-se à modificação dessa variável ao longo do tempo. Por exemplo, em 1995 a variável é construída contendo 10 categorias referentes ao país de origem do estrangeiro, já em 2015, a variável passa a ser construída com 32 categorias.

Nesse sentido, esse artigo tem como objetivo discutir estratégias metodológicas com intuito de contribuir para a superação dessas fragilidades existentes na RAIS e ampliar o conhecimento de mais pesquisadores sobre essa importante fonte de dados, permitindo que o aumento de usuários dessas informações.

### **3. Métodos**

A partir da obtenção dos microdados da RAIS de 1995 a 2015, junto ao Ministério do Trabalho e Emprego são discutidos os procedimentos necessários para realização da análise dos indivíduos no mercado de trabalho, em específicos dos estrangeiros, e superação das fragilidades existentes nessa base de dados.

### **4. Resultado**

Em sua versão enviada pelo MTE, a RAIS mostra a situação de cada trabalhador ao final de cada ano previsto na trajetória, considerando aspectos relacionados ao próprio trabalhador, ao estabelecimento empregador e ao vínculo empregatício que o mesmo possui. Ressalte-se que, ao final de cada ano, para cada trabalhador é armazenado na base apenas o vínculo ativo, isto é, aquele sem data de desligamento ou, para os trabalhadores que não se encontrem em atividade em 31 de dezembro, o vínculo com data de desligamento mais recente. Desse modo, considerando-se um ponto inicial no tempo, acompanham-se os mesmos indivíduos em sucessivos momentos, podendo-se, assim, compor a sua trajetória ocupacional,

em um determinado intervalo de tempo. O acompanhamento da trajetória é feito através do PIS do trabalhador, pesquisado nos registros da RAIS. Assim, os dados referentes a um determinado trabalhador são captados sempre que o mesmo tiver vínculo formal e o seu empregador responder à declaração anual da RAIS. É possível identificar, então, para um determinado intervalo de tempo, aqueles trabalhadores que permanecem empregados ao longo de todo o período, aqueles que, sendo desligados em um ano específico, conseguem retornar ao mercado formal e também aqueles que, após o desligamento, não mais retornam ao mercado formal. Ao incorporar a dimensão das trajetórias ocupacionais, a RAIS traz, inegavelmente, um avanço significativo aos estudos sobre o comportamento do mercado de trabalho.

Nessa perspectiva, a unidade de análise da RAIS é o vínculo empregatício, na medida em que indivíduos que possuem mais de um vínculo de trabalho em um mesmo ano aparecem mais de uma vez no banco de dados naquele respectivo ano. Isso ocorre em casos de: a) pessoas que são empregadas em mais de uma empresa ao mesmo tempo; b) trabalhadores que mudam de emprego durante o ano. Para tanto, o primeiro procedimento metodológico necessário de ser feito na RAIS é a transformação da unidade de análise da RAIS de vínculo empregatício, como é originalmente construída, para indivíduo. Portanto, é preciso identificar os casos de indivíduos com mais de uma informação no banco.

Além do PIS, que é um número que permite identificar cada indivíduo ao longo dos anos, a partir de 2002, a RAIS também apresenta o número do Cadastro Pessoas Físicas (CPF). Porém, como analisado por Cavalcanti *et al* (2015) no Relatório Anual do Observatório das Migrações Internacionais, o PIS é a identificação mais consistente. Dessa forma, utiliza-se como variável identificado o PIS do indivíduo. Em seguida, são realizados os seguintes passos: 1º) Separação dos indivíduos com vínculo ativo em 31/12 do ano analisado; 2º) Para um mesmo PIS, o são escolhidos para permanecer na base a linha que possuir maior número de horas semanais contratadas; 3º) Se houver igualdade no número de horas trabalhadas, então é selecionada a linha com o maior salário no mês de dezembro; 4º) Caso número de horas semanais contratadas e salário no mês de dezembro sejam iguais, é selecionada a linha com o maior tempo de serviço; 5º) Persistindo as igualdades, então é selecionada a linha com maior salário contratual a permanecer na base. A partir deste ponto, se as variáveis eleitas para o filtro permanecerem

iguais, é escolhida a primeira linha empatada. Ressalta-se que a probabilidade de ocorrência de igualdades após o quarto passo é muito pequena.

Utilizamos como variável identificadora de cada indivíduo o PIS e selecionamos apenas as informações referentes ao trabalho principal dos indivíduos, excluindo em média 17% dos casos.

Outra limitação da RAIS refere-se à existência apenas de informações sobre os trabalhadores do mercado formal. Para superação dessa dificuldade propõe-se a inserção de uma variável *predict\_Formal*, como variável a ser utilizada como controle em possíveis estimações multivariadas, criada com intuito de corrigir possíveis vieses de seletividade existentes na amostra. O viés refere-se ao fato da RAIS apenas incluir trabalhadores do mercado formal, entretanto, sabe-se que há um grande número de trabalhadores em ocupações informais no Brasil. A variável *predict\_Formal* que se refere a probabilidade de um trabalhador estar inserido no mercado formal. Para estimação dessa probabilidade, é aplicada uma regressão logística aos dados da RAIS agregados aos dados do Censo demográfico brasileiro para os indivíduos no setor informal. Na regressão foram incluídas as seguintes variáveis independentes: idade, gênero, escolaridade, se é estrangeiro ou brasileiro, se é portador de deficiência e região do país. Dessa forma, é gerado um escore de propensão para todos os indivíduos do banco de dados.

Em relação à variável sobre a nacionalidade dos estrangeiros, que foi modificada ao longo dos anos, tendo suas categorias modificadas de 10 para 32 categorias referentes ao país de origem do estrangeiro, propõe-se a reclassificação das informações com a junção dos bancos de dados dos diferentes anos.

## **5. Referência bibliográfica**

STERNBERG, SHEILA SARA WAGNER. "A RAIS MIGRA como instrumento de análise do mercado de trabalho: um exercício a partir dos dados do Rio Grande do Sul." *Indicadores Econômicos FEE* 28, no. 4 (2001): 183-199.

CAVALCANTI, L. et al. A inserção dos imigrantes no mercado de trabalho brasileiro. Cadernos do Observatório das Migrações Internacionais. Brasília 2014.