



PPgECi

PROGRAMA DE PÓS-GRADUAÇÃO
EM EDUCAÇÃO EM CIÊNCIAS | UFRGS

**UMA MEDIDA DE IMPACTO DE POLÍTICAS PÚBLICAS DE
INVESTIMENTO NA CIÊNCIA: ANÁLISE DE EQUIDADE E O USO DE
ECONOMETRIA COM O MODELO DAS DIFERENÇAS EM DIFERENÇAS
NO UNIVERSAL E EM BOLSAS DE DOUTORADO**

Autor:

ALBERTO DE CAMPOS E SILVA



ORCID.ORG/0000-0001-6506-1988
LATTES.CNPQ.BR/7281587998425548

LINHA DE PESQUISA:

EDUCAÇÃO CIENTÍFICA: PRODUÇÃO CIENTÍFICA E AVALIAÇÃO DE
PRODUTIVIDADE EM CIÊNCIA

ORIENTADOR:

PROF. DR. ALEXANDRE GUILHERME MOTTA SARMENTO

BRASÍLIA, DF
Julho de 2023

Alberto de Campos e Silva

UMA MEDIDA DE IMPACTO DE POLÍTICAS PÚBLICAS DE INVESTIMENTO NA CIÊNCIA: ANÁLISE DE EQUIDADE E O USO DE ECONOMETRIA COM O MODELO DAS DIFERENÇAS EM DIFERENÇAS NO UNIVERSAL E EM BOLSAS DE DOUTORADO

Tese apresentada ao Programa de Pós-graduação em Educação em Ciências (PPgECi) na Universidade Federal do Rio Grande do Sul, como requisito para obtenção do título de doutor.

Orientador: Prof. Dr. Alexandre Guilherme Motta Sarmiento

Brasília

2023

CIP - Catalogação na Publicação

de Campos e Silva, Alberto
UMA MEDIDA DE IMPACTO DE POLÍTICAS PÚBLICAS DE
INVESTIMENTO NA CIÊNCIA: ANÁLISE DE EQUIDADE E O USO
DE ECONOMETRIA COM O MODELO DAS DIFERENÇAS EM
DIFERENÇAS NO UNIVERSAL E EM BOLSAS DE DOUTORADO /
Alberto de Campos e Silva. -- 2023.
173 f.
Orientador: ALEXANDRE GUILHERME MOTTA SARMENTO.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Instituto de Ciências Básicas da Saúde,
Programa de Pós-Graduação em Educação em Ciências:
Química da Vida e Saúde, Porto Alegre, BR-RS, 2023.

1. Avaliação de Políticas Públicas. 2. Mulheres na
Ciência. 3. Indicadores de CT&I. 4. Avaliação de
Impacto. 5. Currículo Lattes. I. GUILHERME MOTTA
SARMENTO, ALEXANDRE, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com
os dados fornecidos pelo(a) autor(a).

Alberto de Campos e Silva

UMA MEDIDA DE IMPACTO DE POLÍTICAS PÚBLICAS DE INVESTIMENTO NA CIÊNCIA: ANÁLISE DE EQUIDADE E O USO DE ECONOMETRIA COM O MODELO DAS DIFERENÇAS EM DIFERENÇAS NO UNIVERSAL E EM BOLSAS DE DOUTORADO

Aprovado em 24 de julho de 2023.

Banca examinadora:

Prof. Dr. Alexandre Guilherme Motta Sarmiento (UFRGS, orientador)

Profa. Dra. Maria do Rocio Fontoura Teixeira (UFRGS, relatora)

Prof. Dr. Diogo Losch de Oliveira (UFRGS)

Profa. Dra. Vanessa Carvalho de Andrade (UNB)

Dedico este trabalho à minha esposa, amada, a beleza em minha jornada. A quem conheceu meu melhor e também o meu pior. A mulher que, vendo a escuridão em mim, mesmo assim, permaneceu enfim. Aquela que me ajudou a tudo superar, o caminho enfrentar e a felicidade encontrar. Aquela que enfrentou a dor e com amor aguentou. Minha, pra sempre minha, a senhora das flores.

Agradecimentos

Primeiramente, a meu amor, meu tudo, minha vida. A mais bela, que tudo me deu e tudo me dá. Sem ela, nada seria, nada teria e nada alcançaria. Minha fundação, meu apoio, minha alegria, que comigo sofreu, que sempre me amou e sempre me apoiou. A ela devo a vida, e a ela dedico toda minha vida. A única capaz de me aceitar como sou, de me amar por quem eu sou e ser feliz por ser quem sou. Também a meus filhos, que ajudaram quando tinham de ajudar e suportaram quando tinham de suportar, mesmo quando podiam apenas surtar.

Meu pai sempre foi meu exemplo, meu guia e mentor. Mesmo já falecido, me acompanha no dia a dia, me orientando, me guiando e me protegendo. Foi graças a ele que me tornei quem sou e a ele devo minha força. À minha mãe, que tanto sacrificou para me criar. Também à minha irmã, Cláudia, o alicerce que mantém minha família em pé, apesar das dificuldades da vida

Ao meu orientador, Dr. Alexandre Motta, excelente professor, dou especial agradecimento. Sem ele, nada disso seria possível. Me aceitou em um momento de dificuldade, tornando-se meu orientador, meu salvador, um amigo e companheiro. Ele sabe dar a liberdade criativa e a orientação. Obrigado, professor, por ter me aceito e orientado.

Ao professor Hilário Alencar da Silva, a quem em hora fundamental, me aconselhou. A ele também devo esse título de doutor, pois, se não fosse sua ajuda nessa hora tão crítica, não conseguiria trilhar na direção correta e teria me perdido no labirinto da vida. Às minhas antiga e atual chefias no CNPq, Raquel e Renta, pela paciência nos momentos em que precisei.

Também não posso deixar de dar um sincero agradecimento aos professores Dra. Maria do Rocio Fontoura Teixeira, Dr. Diogo Losch de Oliveira e Dra. Vanessa Carvalho de Andrade, que realizaram leitura atenta deste texto e deram construtivas contribuições e críticas. Sem as importantes contribuições, o texto nunca seria tão completo e dinâmico. Agradeço também à minha colega, Juliana Martins Ferreira, que tão prontamente realizou importantes revisões dos textos.

Finalmente, a todos que diretamente ou indiretamente me ajudaram. Mais do que nunca vejo que o caminho até aqui não foi fácil, mas que somente consegui trilhar este tortuoso e ardiloso caminho, vencendo pântanos e areias movediças, apenas graças aos ombros dos gigantes que me carregaram, apoiaram e guiaram.

RESUMO

O objetivo desta pesquisa é desenvolver método para calcular o Impacto de uma Política Pública de Financiamento em Ciência. Como diretriz, a metodologia foi concebida de forma que seja escalável a várias políticas públicas, transparente, que permita a comparação entre avaliações diferentes, tanto de forma longitudinal (entre políticas) quanto vertical (no tempo) e que seja adequada a usos frequentes e automatizados ou semiautomatizados. O método infere e mede a causalidade entre um dos diversos indicadores retirados do Currículo Lattes, comparando esse indicador entre duas distintas datas e levando-se em consideração o investimento realizado. Para isso, foi desenvolvido modelo quase-experimental de análise econométrica baseada no método das Diferenças em Diferenças, método recomendado pelo Banco Mundial, com o uso de múltiplas regressões lineares multivariadas. Para colocar o método em prática, foram baixados e tratados Currículos Lattes dos bolsistas do recorte, com o uso da linguagem Python e do banco de dados PostgreSQL, de modo que cada atributo em que “ANO” estivesse disponível, foi retirado um indicador, como por exemplo “Artigo Publicado” e “Orientação de Doutorado Concluída”.

Para testar o método proposto, ele foi aplicado em duas distintas políticas públicas de investimento à ciência e os resultados estudados. No entanto, percebeu-se nesse processo a possibilidade de presença de misoginia no nas concessões do CNPq antes nunca detectada ou estudada. Dessa forma, essa misoginia também foi estudada, encontrando-se forte indicador de presença de discriminação sexual contra as mulheres, presente principalmente no valor financiado a menor para mulheres e no teto de vidro¹ para cargos de alto *status* ou poder.

Encontrou-se também que as principais agências de fomento federais (a CAPES e o CNPq) possuem impactos quase equivalentes em suas políticas de concessão de bolsas de doutorado e que o financiamento na Chamada Universal possui melhor custo-efetividade. Comprovou-se, portanto, que a ferramenta pode ser aplicada e que traz medida do impacto do investimento público em ciência, podendo ser amplamente adotada e até mesmo automatizada.

Palavras-chave: Indicadores de C&T, Avaliação de Impacto, Formação de Recursos Humanos, Quase-Experimento, Currículo Lattes, Política Pública de Investimento em Ciência, Discriminação Sexual, Igualdade de Gênero, Mulheres na Ciência, Avaliação de Políticas Públicas, Mulheres na CTEM, CNPq.

¹ “O teto de vidro (glass ceiling) é um fenômeno social que a partir de barreiras culturais organizacionais, familiares e individuais, dificulta o acesso das mulheres a posições de liderança, principalmente aos mais altos níveis na hierarquia organizacional.” Carneiro, Luziberto Barrozo, In: *Teto de vidro: um estudo sobre os fatores deste fenômeno no Brasil sob a percepção das mulheres gestoras*. <http://repositorio.furg.br/handle/1/7886>

ABSTRACT

The objective is to present a method to calculate the Impact of a Public Policy on Science Funding. As a guideline, the methodology was conceived in such a way that it is scalable to several public policies, transparent, allows comparison between different evaluations, both longitudinally (between policies) and vertically (in time) and adequate to frequent and automated or semi-automated uses. The method infers and measures the causality between one of the several indicators present in the Lattes Curriculum, comparing this indicator between two different dates and considering the investment made. For this, a quasi-experimental econometric model was developed based on the Differences in Differences method (which was recommended by the World Bank) with the use of multiple multivariate linear regressions. To put the method into practice, Lattes Curriculum were downloaded and treated, using the Python language and the PostgreSQL database, so that each attribute in which “YEAR” was available resulted in an indicator (such as “Article Published” and “Concluded Doctoral Orientation”).

To test the proposed method, it was applied in two different public policies for investment in science and the results studied. However it was noticed in this process the possibility of misogyny in CNPq’s funding that had never been detected or studied before. Thus, this misogyny was also studied, finding a strong indicator of the presence of sexual discrimination against women mainly present in the amount financed at minus for women and in the glass ceiling for positions of high status or power.

It was also found that the main federal funding agencies (CAPES and CNPq) have almost equivalent impacts on their policies for awarding doctoral scholarships, but that funding at Universal has an even greater impact. It was proved that the tool can be applied and that it measures the impact of public investment in science and can be widely adopted and even automated.

Keywords: S&T indicators, Sexual Discrimination, Gender Equality, Women in Science, Evaluation of Public Policies, Women in STEM, CNPq, Impact Assessment, Human Resources Training, Quasi-Experiment, Curriculum Lattes, Public Policy for Investment in Science.

LISTA DE TABELAS

Tabela 1 – Proporção entre Homens e Mulheres no Universal 2014	27
Tabela 2 – Proporção entre Homens e Mulheres no Universal 2014	42
Tabela 3 – Impacto de Concessão de Bolsas de Doutorado pelo CNPq e pela CAPES.	91
Tabela 4 – Custo-Efetividade da concessão e Bolsas a Doutorandos 2014 por concessão ou não e por valor investido.	92
Tabela 5 – Impacto da Chamada Universal.....	99
Tabela 6 – Impacto de Concessão de Bolsas de Doutorado pelo CNPq e pela CAPES.	105
Tabela 7 – Comparação da Custo-Efetividade no ano de 2014: Concessão de Bolsas de Doutorado e Universal.....	106
Tabela 8 – Regressão Linear com Medida de um Impacto de Concessão de Bolsas de Doutorado	146
Tabela 9 – Impacto de Concessão de Bolsas de Doutorado pelo CNPq e pela CAPES.	147
Tabela 10 - Diferença em Diferenças por Pareamento dos Doutorandos 2014 por soma simples.	150
Tabela 11 - Impacto de Bolsas do CNPq aos Doutorandos 2014.....	152
Tabela 12 - Custo-efetividade de bolsas de doutorado pelo CNPq a quem iniciou seu doutorado em 2014.	153
Tabela 13 – Custo-Efetividade da concessão e Bolsas a Doutorandos 2014 por concessão ou não e por valor investido.....	153
Tabela 14 – Impacto da Chamada Universal 2014 em Formação Humana a nível de Pós Graduação	168
Tabela 15 – Impacto da Chamada Universal.....	170

LISTA DE FIGURAS

Figura 1 - A Ciência Aberta (Open Science) de acordo com a Comissão Europeia	30
Figura 2 – Mulheres no CNPq.....	44
Figura 3 - Mulheres no CNPq	45
Figura 4 - Porcentagem de Mulheres na Ciência.....	53
Figura 5 - Participação das Mulheres nas Ações do CNPq ao longo do tempo	54
Figura 6 – Porcentagem de Mulheres contempladas no CNPq, por Faixa de Valores.....	55
Figura 7 - Porcentagem de Recursos Concedidos à Mulheres no CNPq, por faixa de valores	55
Figura 8 - Delta, por faixa de valores	56
Figura 9 - Mulheres no CNPq por Área - 2021	57
Figura 10 - Participação das Mulheres - Saúde e Exatas - 2021	57
Figura 11 - Participação das Mulheres no CNPq, por Linha de Fomento - 2002 a 2021.....	58
Figura 12 - Participação das Mulheres em APQ no CNPq, por ano.	58
Figura 13 - CNPq - Mulheres e Bolsas Produtividade por ano	60
Figura 14 – Fluxo do Fomento do CNPq	72
Figura 15 – Fluxo de Análise e Julgamento das Propostas	73
Figura 16 – Valor Total Concedido pelo CNPq para Bolsas GD	76
Figura 17 – Valor financiado em Chamadas Universais	77
Figura 18 – Nuvem de Palavras envolvendo Impacto na Ciência.....	81
Figura 19 – A produção do conhecimento segundo a visão representativa dos dados.....	86
Figura 20 – Processo de Ciência de Dados.....	86
Figura 21 – Diagrama de Venn desenvolvido por Conway.....	87
Figura 22 – Os sete “Vs” do Big Data.....	87
Figura 23 – Representações da Tripla Hélice em termos de um Diagrama de Venn ou como triângulo.....	122
Figura 24 – A Hélice Tripla como um sistema cartesiano de Coordenadas.....	122
Figura 25 – Ferramentas de Análise da Pesquisa pertencentes ao Grupo 1 e 2	123
Figura 26 – O método da Diferença em Diferenças	128
Figura 27 - Distribuição da concessão de bolsas de doutorado daqueles que iniciaram seu doutorado em 2014.	141
Figura 28 – Distribuição do indicador Artigos Publicados de acordo com a concessão de bolsas	141
Figura 29 – Distribuição dos Erros da Regressão Linear do Indicador Artigos Publicados ..	142

Figura 30- Ano de conclusão do doutorado iniciado em 2014 (%)..... 143

Figura 31 - Comparação do tempo de doutoramento - com e sem bolsas de estudo..... 144

LISTA DE EQUAÇÕES

Equação 1 – Elasticidade dos Recursos Femininos no Universal 2014	43
Equação 2 - Elasticidade.....	44
Equação 3 - Elasticidade.....	52
Equação 4 – Modelo Econométrico para o Cálculo do Impacto.	83
Equação 5 – Medida de um Impacto por meio de Regressão multivariada categórica.....	91
Equação 6 – Medida de Impacto por meio de Regressão multivariada categórica para o CNPq:	92
Equação 7 – Regressão Linear entre índice atual, índice anterior e financiamento.	94
Equação 8 - Regressão Linear entre índice atual, índice anterior e financiamento.....	99
Equação 9 – Modelo Matemático das Diferença em Diferenças.....	129
Equação 10 – Regressão linear em Diferença em Diferenças	129
Equação 11 – Modelo Econométrico para o Cálculo do Impacto.	130
Equação 12 – Medida de um Impacto por meio de Regressão multivariada categórica.....	137
Equação 13 – Medida de Impacto por meio de Regressão multivariada categórica para o CNPq:	139
Equação 14 - Financiamento influencia no tempo de doutoramento?.....	144
Equação 15 – Medida de um Impacto por meio de Regressão multivariada categórica.....	145
Equação 16 - Impacto Negativo em Novos Vínculos.....	149
Equação 17 - Diferença em Diferenças	150
Equação 18 – Cálculo do Impacto pelo método da Diferença em Diferenças	150
Equação 19 – Medida do Impacto do CNPq por meio de Regressão multivariada.....	151
Equação 20 – Modelo de Regressão Linear para comparar a Produtividade no Tempo por meio de um Estudo Bibliométrico (MILESI et al., 2014).	162
Equação 21 – Modelo de Regressão Linear sem variáveis temporais.....	163
Equação 22 – Modelo de Regressão Linear simplificada.....	163
Equação 23 - Regressão Linear entre índice atual, índice anterior e financiamento.....	163

SIGLAS E ABREVIATURAS

Sigla/Abreviatura	Nome
APQ	Auxílios à Pesquisa
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CT&I	Ciência, Tecnologia e Inovação
CTEM	Ciência, Tecnologias, Engenharia e Matemática
DT	Bolsas de Desenvolvimento Tecnológico
DTI	Bolsas de Desenvolvimento Tecnológico e Industrial
LDB	Lei de Diretrizes e Bases da Educação Nacional
MCTI	Ministério da Ciência, Tecnologia e Inovações
OECD	Organização para a Cooperação e Desenvolvimento Econômico
P&D	Pesquisa e Desenvolvimento
PNE	Plano Nacional de Educação
PQ	Bolsas Produtividade em Pesquisa
SQL	Structured Query Language (Linguagem de Consulta Estruturada)
STEM	Science, Technology, Engineering, Mathematics
TI	Tecnologia da Informação
Universal	Ações ou Chamadas UNIVERSAL no CNPq, ou CHAMADA UNIVERSAL MCTI/CNPQ N° 14/2014.

SUMÁRIO

1.	PARTE I - INTRODUÇÃO	20
1.1.	Investimento público em P&D	20
1.2.	Avaliação do Investimento.....	22
1.3.	Organização desta pesquisa.....	24
1.4.	Classificação desta pesquisa	25
1.5.	Segunda Parte - Equidade	26
1.5.1.	Introdução.....	26
1.5.2.	Hipótese.....	27
1.5.3.	Objetivo	27
1.5.4.	Conclusão	28
1.6.	Terceira Parte - Análise do Impacto.....	28
1.6.1.	Hipótese.....	29
1.6.2.	Objetivos	29
1.6.3.	Conclusão	29
2.	ESTA PESQUISA NO CONTEXTO DA CIÊNCIA ABERTA.....	30
2.1.	Coleta de Dados	32
2.1.1.	Currículo Lattes	32
2.1.2.	Pagamentos Realizados	33
2.1.3.	Demanda Bruta.....	33
2.1.4.	Sexo.....	33
2.1.5.	Coleta de dados de financiamento por outras agências.....	34
2.1.6.	Período da análise.....	34
3.	PARTE II - ANÁLISE DE EQUIDADE: MULHERES NA CIÊNCIA.....	36
3.1.	Referencial Teórico	36
3.2.	Metodologia.....	41
3.2.1.	Delta – Uma medida de Elasticidade	41
4.	ARTIGO: MULHERES NO CNPQ – UMA ANÁLISE DE POSSÍVEIS MECANISMOS DE DESIGUALDADE DE GÊNERO NO FOMENTO FEDERAL À CT&I	47
4.1.	Resumo	48
4.2.	Introdução.....	49
4.3.	O CNPq	50

4.4.	Coleta dos dados	51
4.4.1.	Delta	52
4.5.	A mulher no CNPq.....	52
4.5.1.	O Financiamento às Mulheres ao Longo dos Anos.....	53
4.5.2.	Mulheres no CNPq por Faixa de Valores.....	54
4.5.3.	Mulheres no CNPq por Área do Conhecimento.....	56
4.5.4.	Presença das Mulheres por Linha de Fomento.....	58
4.5.5.	Apoio a Projetos de Pesquisas (APQ)	58
4.5.6.	Bolsas Produtividade (PQ)	59
4.6.	Conclusão	60
5.	PARTE III - ANÁLISE DO IMPACTO DA POLÍTICA PÚBLICA DE INVESTIMENTO NA CIÊNCIA	62
5.1.	Definição do Impacto da Ciência	62
5.2.	A Análise do Impacto da Ciência no Mundo	63
5.2.1.	Estados Unidos	63
5.2.2.	Austrália	64
5.2.3.	Países Baixos.....	65
5.2.4.	Comissão Europeia (E3M)	66
5.2.5.	Alemanha.....	66
5.2.6.	Brasil – MCTI	67
5.2.7.	Brasil – CGEE.....	69
5.2.8.	Brasil – CNPq.....	70
5.2.9.	Conclusão	73
5.3.	Escolha das Políticas Públicas a serem analisadas	74
5.3.1.	Bolsas de Doutorado	76
5.3.2.	A Chamada Universal	77
5.3.3.	Conclusão	78
5.4.	Método de levantamento bibliográfico	78
6.	METODOLOGIA	83
7.	CONSIDERAÇÕES ADICIONAIS NESTA PESQUISA	85
7.1.	Ciência de Dados	85
7.2.	O Viés do Duplo Financiamento	88
7.3.	A dimensão da Qualidade e a falta de análise dessa dimensão neste estudo.	90
8.	O IMPACTO DE BOLSAS DE DOUTORADO	91

9.	O MODELO ECONOMÉTRICO E A REDUÇÃO DE DIMENSÕES PROPOSTAS.....	94
10.	O IMPACTO DO UNIVERSAL 2014.....	99
11.	PARTE IV - CONCLUSÃO	102
11.1.	Conclusões sobre a Análise de Equidade – Misoginia no CNPq.....	102
11.2.	Conclusões sobre o Impacto de Políticas Públicas em Ciências.....	104
11.2.1.	Comparação entre o impacto da concessão de bolsas de doutorado entre a CAPES e o CNPq;	104
11.2.2.	Comparação entre o impacto da concessão de bolsas de doutorado e no investimento no Universal 2014, ambos realizados pelo CNPq.	106
11.2.3.	Terceira missão da universidade, o impacto social da ciência.	108
11.3.	Considerações Sobre o Cálculo do Impacto.....	109
11.4.	Limitações e Estudos Futuros	110
11.5.	Considerações Finais.....	111
12.	REFERÊNCIAS BIBLIOGRÁFICAS	113
13.	ANEXO I - ARTIGO: INFERINDO A CAUSALIDADE ENTRE FINANCIAMENTO EM CIÊNCIAS E A FORMAÇÃO DE RECURSOS HUMANOS: UM MODELO ECONOMÉTRICO PARA MEDIR O IMPACTO DE POLÍTICAS PÚBLICAS POR UM QUASE-EXPERIMENTO.....	120
13.1.	Introdução.....	120
13.2.	Revisão Bibliográfica	123
13.2.1.	Ferramentas para a Análise do Impacto	123
13.2.2.	Indicadores para a Análise Econométrica	124
13.2.3.	O Currículo Lattes e Data Mining	125
13.2.4.	Cientometria	126
13.2.5.	Análise Econométrica.....	126
13.2.6.	Método das Diferença em Diferenças	127
13.3.	Modelo Econométrico – Regressão Linear	130
13.4.	Conclusão	132
14.	ANEXO II – ARTIGO: O IMPACTO DE BOLSAS DE DOUTORADO NA PÓS-GRADUAÇÃO BRASILEIRA: UMA ANÁLISE PELO CURRÍCULO LATTES DO CNPQ.....	133
14.1.	Introdução.....	134

14.2.	Metodologia.....	136
14.3.	Análise dos dados.....	140
14.3.1.	Distribuição dos Financiamentos	140
14.3.2.	Distribuição dos Dados	141
14.3.3.	Ano de Conclusão do Doutorado	143
14.3.4.	Impacto do Financiamento	145
14.3.5.	Artigos Publicados	148
14.3.6.	Vínculos.....	148
14.3.7.	Pareamento	149
14.3.8.	Análise Custo-Efetividade.....	152
14.4.	Conclusão	154
15.	ANEXO III – ARTIGO: O IMPACTO DE UM INVESTIMENTO PÚBLICO EM CIÊNCIA NA FORMAÇÃO DE RECURSOS HUMANOS EM NÍVEL DE MESTRADO E DOUTORADO: O CASO DO UNIVERSAL 2014....	156
15.1.	Introdução.....	157
15.2.	Metodologia.....	160
15.3.	Currículo Lattes	164
15.4.	Obtenção e tratamento dos dados.....	165
15.5.	Resultados e discussões	167
15.6.	Considerações finais	170

Apresentação

Se fosse citar aquilo que é de mais importante em minha vida, sou casado com uma mulher maravilhosa e tenho três lindos filhos. Quase tudo o que faço, direta ou indiretamente, é por eles. Mas, estamos aqui para falar de minha vida profissional.

Atualmente, sou Analista de Ciência e Tecnologia no CNPq, lotado no SEADM-DCTI, prestando assessoria à Diretoria Científica Adjunta. Mas, para chegar aqui, não foi fácil. Tenho 46 anos, sou formado em física pela UnB, tendo dado aulas (de física) para o ensino fundamental por 2 anos (concursado GDF). Como físico aprendi a sempre procurar como e por que as coisas funcionam, a ver os motivos por trás do que vemos, a procurar a verdade por trás da aparência. Como professor, vi a importância de aprender e de ensinar. Aprendi a estudar e a colocar esse conhecimento em prática.

Me mudei para Goiânia e por lá trabalhei na justiça, por 9 anos, aprendendo muito sobre direito. Grande parte de uma ampla base jurídica que possuo foi devido a esse trabalho na Justiça de Goiânia. Essa base jurídica mudou minha forma de ver o direito, a justiça e o governo. Como era encarregado inclusive da parte de informática, e como sempre gostei de programar, também aprendi programação de banco de dados (SQL), web-design, C#, criação de máquinas virtuais, desenvolvimento de sistemas, *Active Directory* e manutenção de micros. Assim, fui gerente, administrador, advogado e programador (sou graduado apenas em física).

Um pouco ambicioso, comecei a me preparar para concursos públicos no intuito de melhorar minha qualidade de vida, fazendo, para isso, pós-graduação em gestão pública. Depois de três anos de estudos árduos, passei em alguns concursos, tendo escolhido o CNPq para trabalhar, não só por ter afinidade ao cargo, como também pelo bom salário.

Para alavancar minha carreira, fiz mestrado em Gestão Econômica do Meio Ambiente, aprendendo muito sobre economia. O mestrado também mudou minha forma de ver o mundo, ver o papel do estado e da economia, formando a minha base econômica. Nele, aprendi quão volátil e imaterial é o dinheiro, o papel do governo e como ele funciona, as limitações e forças das políticas públicas, a diferença entre custo e valor, sobre as comunas, enfim, muito do porquê políticas públicas existem, como e quando elas são importantes e como realizar avaliações.

Após o mestrado o próximo passo foi procurar um doutorado. Como sempre tive afinidade com a educação, mas, ao mesmo tempo, por ter formação interdisciplinar, o doutorado no Programa de Pós-Graduação em Educação em Ciências (PPgECi) caiu como uma luva. Embora na área de Educação em Ciências, um de seus temas é a avaliação das ciências, objeto deste trabalho.

Autodidata e com gosto por conhecimento, ainda tenho especializações em Gestão Governamental, Projetos, Processos, Propriedade Intelectual, Impacto de Políticas Públicas, Plano Estratégico, Banco de Dados, Avaliação de Ações e Avaliação de CT&I. Tenho experiência com programação (especialmente Python) e banco de dados relacional, além de web-design. Ainda tenho conhecimentos de cinofilia (por 10 anos tive criação de cães), fotografia (ainda hoje sou fotógrafo) e ensino (afinal, essa foi minha formação).

Ou seja, minha trajetória me ensinou um pouco de tudo. Física, direito, gestão pública, economia, arte, programação, gerência e até contabilidade. Tendo administrado um cartório por vários anos, não podia deixar de aprender sobre a importância da legalidade e da gestão de riscos. Aprendi a não recuar diante de uma necessidade, a encontrar procurar a melhor solução possível, mesmo não sendo aquela ideal.

Foi por isso que, diante da necessidade do meu atual local de trabalho, o CNPq, de provar ao mundo a importância da ciência e o benefício que ele traz à sociedade, coloquei tudo o que aprendi em funcionamento para alcançar esse objetivo. Para isso, e aproveitando minha vocação autodidata, realizei cursos em Avaliação de Políticas Públicas, atingindo certificação pela ENAP. Finalmente, observando falta de conhecimento em *Big Data*, realizei MBA no assunto.

Assim, unindo toda a minha experiência, usei o que aprendi em programação, direito, economia, avaliação de políticas públicas e desenho de processos para encontrar uma forma de analisar os efeitos dos investimentos realizados pelo CNPq. Ao menos, essa foi minha intenção, trabalho ao qual me dediquei por anos de minha vida.

O resultado segue nesta tese, que é interdisciplinar por natureza. Afinal, ela envolve políticas públicas, educação, economia, programação, *big data*, ciência de dados e direito, sendo que, não fosse minha formação em física e a forma como aprendi a buscar as relações e a forma de funcionamento do mundo, nada disso seria possível. Mas, minha intenção não é parar por aqui, mas aprofundar ainda mais o estudo e o implementar oficialmente. Além disso, desenvolver um programa e um manual capaz de permitir que muitos outros estudos sejam realizados.

Finalmente, gostaria de dizer que escrever essa tese foi especialmente difícil. Talvez porque o estudo envolva tantas áreas do conhecimento, que se interrelacionam e se emaranham, talvez porque eu mesmo não consiga pensar de forma direta e linear. De qualquer forma, foi um desafio escrever um texto didático, capaz de levar o leitor pelas diversas áreas e até o resultado final.

Costumo pensar que, no conhecimento, o início, o meio e o fim estão ligados. Normalmente, não há causa e efeito, mas uma causa que costuma ser um efeito e um efeito que se torna uma causa. Para explicar o mundo, todos os saberes se unem naturalmente, como milhares de peças de um quebra-cabeças, que pode ser montado em qualquer ordem. Pior, diferentemente de um quebra-cabeças, cada peça pode estar em múltiplos lugares ao mesmo tempo, num mapa multidimensional. Não há um caminho linear que leve a um novo conhecimento, mas esse sempre é dependente de muitas ligações cognitivas, nem sempre fáceis de explicar. Assim, espero ter conseguido, mesmo que parcialmente!

Espero uma boa leitura e que Deus os tenha!

1. PARTE I - INTRODUÇÃO

1.1. Investimento público em P&D

Investimento público em P&D pode levar ao desenvolvimento de novas tecnologias, em aumento do bem-estar social e é um dos fatores que interferem diretamente na produção científica e na formação de recursos humanos. De acordo com Salter et al. (2001), são seis as contribuições da pesquisa financiada pelo setor público: (1) expansão do estoque do conhecimento; (2) formação de recursos humanos; (3) criação de novos instrumentos e metodologias; (4) formação de redes e estímulo a interações sociais; (5) aumento da capacidade de solução de problemas técnico/científicos; e (6) criação de novas firmas.

Destaca-se que as seis categorias listadas não são dependentes entre si, mas possuem áreas comuns, se correlacionam e se influenciam e que, no geral, os estudos permitem concluir que os benefícios econômicos do financiamento público à pesquisa: são reais e substanciais, embora de difícil mensuração; se realizam de várias formas diferentes e dependentes do país/área/pesquisa; que a principal questão é como melhor organizar o sistema de financiamento público para trazer o maior benefício possível (SALTER; MARTIN, 2001).

É por isso que muitos países adotaram novas abordagens, baseadas na excelência da pesquisa, na forma como os investimentos públicos são realizados. Reino Unido, Austrália, Países Baixos e a Comissão Europeia, além de vários outros países, implementaram sistemas de avaliação dos investimentos em P&D como forma de oferecer suporte à decisão do financiamento realizado (E3M, 2012; MELOROSE; PERROY; CAREAS, 2015; VSNU; NWO; KNAW, 2015). Uma das principais formas de investimento pelo estado em CT&I, tanto no Brasil quanto no mundo, é por meio de subsídios.

Subsídios são como impostos negativos (PINDYCK; RUBINFELD, 2010) ou incentivos financeiros. A OCDE define subsídios como medidas governamentais para a redução dos preços para abaixo do nível de mercado para consumidores ou acima para produtores. Uma outra forma de ver o subsídio é a redução dos custos aos consumidores ou aos produtores, direta ou indiretamente (DOUGLAS, 1996).

Muitas atividades beneficiam a sociedade sem o respectivo retorno econômico, de modo que o investimento pelo setor público para incentivar a prática dessas atividades é uma medida desejável. De acordo com Parry (1997), Subsídios Pigouvianos são aqueles que incentivam as externalidades positivas no valor exato da diferença do benefício da atividade à sociedade com o valor gasto/recebido e são, portanto, eficientes e devem ser aplicados. Assim,

exemplificando, o salário do pesquisador ativo é, em média, inferior a seu valor junto à sociedade, de modo que o subsídio se torna economicamente eficiente e desejável.

Como exemplos de atividades com externalidades positivas temos a saúde, a educação e o saneamento básico. Um exemplo simples seria a instalação de um parque público (ou qualquer outra benfeitoria), em que todos os que estão lá se beneficiam, mas que quem construiu nada recebe por isso. Um belo jardim, atividades sociais ou o “fazer o bem sem ver a quem” também são exemplos. Atividades de Educação, Ciência, Tecnologia e Inovação (CT&I), bem como de Pesquisa e Desenvolvimento (P&D), são exemplos de bens meritórios com externalidade positivas que devem receber subsídios Pigouvianos, principalmente quando não realizadas por empresas particulares. Afinal, a publicação de um artigo científico ajuda toda a comunidade científica, mas o pesquisador normalmente nada recebe por isso. Em outras palavras, o salário do pesquisador ativo é, em média, inferior a seu valor junto à sociedade, de modo que o subsídio se torna economicamente eficiente e desejável.

No entanto, subsídios podem ser ineficientes e por isso são considerados como uma opção *second best*. Dessa forma, subsídios somente devem ser adotados se não houver alternativa mais eficiente (RIGONATTO, 2006).

Entre as ineficiências do subsídio estão a falta de adicionalidade (paga-se por atividades que teriam sido conduzidas de qualquer maneira), vazamento (parcelas da população não são contempladas ou simples deslocamento das externalidades negativas aos não subsidiados) e incentivos perversos (o incentivo possui resultado não pretendido e prejudicial).

(SILVA, 2018)

Assim, o subsídio é corretamente aplicado e eficiente se internalizar os benefícios externos, de tal forma a recompensar no exato valor dessas externalidades positivas, a todos os que praticarem as atividades que geram esses benefícios. No caso da Política Pública de incentivo à ciência, o valor correto do subsídio seria aquele que corresponde ao benefício social da pesquisa, ou seja, um “imposto negativo” (subsídio, bolsa, financiamento) igual ao valor, definido pela sociedade, de tudo aquilo que decorrer da pesquisa (GOULDER; PARRY, 2008).

Esse valor é de difícil (ou impossível) mensuração, especialmente quando se fala de ciência básica. Ainda assim, torna-se necessária uma avaliação desse investimento, em busca de se ele está de acordo com a externalidade positiva gerada, se há falhas em sua implementação como incentivo perverso, falta de adicionalidade ou mesmo vazamento.

1.2. Avaliação do Investimento

É nesse sentido que se torna cada vez mais essencial que avaliações dos investimentos efetuados sejam realizadas. De acordo com Silva (2015), analisar uma política pública é analisar (OECD, 1997, *apud* SILVA, 2015):

(1) Equidade: A análise da equidade busca verificar quais grupos minoritários ou sociais estão sendo favorecidos ou desfavorecidos (SILVA, 2015). Nesta pesquisa, o grupo social do sexo feminino foi eleito para uma análise de equidade, mas outras opções seriam a racial e a de regionalidade.

(2) Eficiência & (3) Eficácia: Analisar a eficácia de um investimento científico é analisar se ele cumpriu seu objetivo ou meta. Analisar a eficiência é analisar se foi obtido o máximo retorno ao menor custo (o melhor custo/benefício possível). Assim, enquanto a eficácia estaria relacionada à medida do impacto, o benefício estaria mais relacionado à valoração desse impacto. Em ambos os casos, é necessária uma análise do impacto da ciência, que será objeto desta pesquisa.

(4) Incentivo ao esforço máximo ou motivação: É a análise que busca averiguar se a política pública de fato incentiva que os objetivos almejados sejam procurados, sem desvios de finalidade. É procurar saber se quem recebe o benefício de fato busca atingir aquele objetivo. No caso desta pesquisa, medir o impacto da política indiretamente também mede o referido incentivo, principalmente porque, como veremos, o impacto é medido com o uso de diversos indicadores.

(5) Custo administrativo: É o custo governamental para a aplicação da política pública. É o gasto necessário para fazer a máquina pública girar. No caso desta pesquisa, o custo administrativo é simplificado e será equivalente ao investimento realizado.

(6) Aceitação Política: Refere-se à aceitação das políticas pela sociedade (BAUMOL e OATES, 1979, *apud* (SILVA, 2015) e pode normalmente ser verificada por questionários. Busca averiguar a aprovação da política pública pela sociedade.

(7) Permanência: Busca verificar se, mesmo após o encerramento dos investimentos pelo estado, os efeitos almejados permanecem no tempo, se há alteração na cultura ou na sociedade de modo permanente. É verificar se estamos dando o peixe ou se estamos ensinando a pescar.

Ocorre que a análise do retorno do investimento em ciência é especialmente difícil, isso porque o impacto da ciência se dá a longo e imprevisível prazo, além de ser bem público indivisível:

- 1) Os efeitos benéficos de uma pesquisa científica podem surgir apenas anos ou séculos após a realização da mesma (COLWELL et al., 2012).
- 2) Os efeitos científicos são imprevisíveis, ou seja, não há como prever, ao se realizar uma determinada pesquisa, quais serão seus efeitos decorrentes. Dessa forma, os custos e os benefícios futuros de uma determinada pesquisa também são imprevisíveis (COLWELL et al., 2012).
- 3) Cada produto (ou tecnologia) depende de vários conhecimentos científicos distintos e não há como separar o valor econômico de cada conhecimento agregado em um determinado produto final (COLWELL et al., 2012).

Para Simonis et al. (2001), efetividade, custo social, eficiência, justiça distributiva são critérios de primeira ordem, enquanto aceitação política, mudança na competitividade, incentivos seriam critérios de segunda ordem. Critérios estratégicos seriam aqueles relacionados à estrutura e aos objetivos do governo, da sociedade e da cultura (SIMONIS et al., 2001). De acordo com os autores, observa-se falta, na literatura, de avaliações de primeira ordem do investimento em Ciência e Tecnologia realizados pelo governo brasileiro.

De acordo com (GERTLER et al., 2018), as avaliações de impacto são um tipo particular de avaliação que procura o efeito causal de um programa sobre um resultado de interesse. O foco reside apenas no impacto, ou seja, nas mudanças diretamente atribuíveis ao investimento realizado. A avaliação é de especial interesse para uso de autoridades governamentais, gestores de programas, membros da sociedade civil e outras partes interessadas e pode subsidiar o aumento ou a redução do investimento realizado.

Assim, considerando a necessidade dos investimentos públicos em P&D e a forma como são aplicados, este estudo busca avaliar o investimento do País, buscando exemplos abrangentes e buscando uma medida do retorno social do investimento realizado e, mais especificamente, uma medida de impacto. Em especial, busca-se a relação entre o financiamento público em pesquisa e sua interferência na produção científica e na formação de recursos humanos, medindo o retorno do investimento por meio de indicadores, ou seja, uma análise de primeira ordem. Além disso, buscou-se outra análise de primeira ordem, a da equidade (presença de misoginia) na política pública aplicada.

Assim, esta pesquisa avança na questão de Avaliações de Políticas Públicas, principalmente relacionadas ao subsídio à ciência, analisando a Equidade (1), verificando-se a presença de misoginia, e a Eficácia (2) dos recursos concedidos, por meio de uma medida de impacto. O Custo administrativo (5) é também medido de forma simplificada, chegando-se a uma análise custo-efetividade. O (4) Incentivo ao esforço máximo também é indiretamente

medido, uma vez que se verifica se a política pública teve a eficácia esperada ou não. Finalmente, alguns indicadores podem representar a (7) Permanência, como, por exemplo, a formação de doutores, que possui efeito permanente no tempo.

1.3. Organização desta pesquisa

Esta tese se separa em quatro distintas partes. Na primeira, além desta introdução, é apresentado como esta pesquisa se insere no contexto de dados abertos. Dessa forma, o Capítulo 2 indica como os dados foram obtidos, quais são de acesso público, e como obter os dados necessários para que a pesquisa seja refeita.

Na segunda parte é apresentada a análise da equidade. Assim, o Capítulo 3 apresenta o referencial teórico e a metodologia dessa análise, que é realizada no Capítulo 4, que apresenta o Artigo: Mulheres no CNPq – Uma análise de possíveis mecanismos de desigualdade de gênero no fomento federal à CT&I.

A terceira parte trata da análise do impacto do subsídio à pesquisa e se encontra nos capítulos 5 a 11. O Capítulo 5 introduz o assunto e apresenta revisão bibliográfica sobre o assunto. Nele são indicadas as boas práticas de análise da pesquisa e como alguns países realizam essa análise.

O Capítulo 6 apresenta faz breve relato do Anexo I, que apresenta o modelo econométrico no Artigo: Inferindo a Causalidade entre Financiamento em Ciências e a Formação de Recursos Humanos: um Modelo Econométrico para Medir o Impacto de Políticas Públicas por um Quase-Experimento.

O Capítulo 7 realiza considerações adicionais necessárias para aplicar o modelo econométrico em uma política pública de subsídio à pesquisa, de modo que os próximos capítulos realizam essa aplicação na prática em busca de se medir o impacto em duas formas distintas de investimento: a primeira em bolsas e a segunda diretamente em auxílio a pesquisas.

A primeira análise de impacto é realizada no Anexo II, com breve relato presente no Capítulo 8, que apresenta O Impacto de Bolsas de Doutorado. Nele, o impacto da concessão de bolsas por duas distintas agências de fomento (CAPES e CNPq) é medido.

O Capítulo 9 introduz a questão da redução de dimensões e busca explicar os motivos pelos quais o modelo econométrico indicado não apresenta mais variáveis independentes. O motivo de estar presente nesta posição é que sua compreensão depende do artigo anteriormente citado e, por isso, seria de difícil entendimento em capítulos anteriores, sendo uma questão técnica importante para a compreensão da aplicabilidade do modelo.

A segunda análise de impacto é realizada no Anexo III, com breve relato presente no Capítulo 10, que apresenta O Impacto do Universal 2014. Dessa forma, mede-se o impacto do subsídio direto à pesquisa, em que o CNPq concede uma verba a ser livremente gasta na pesquisa conforme projeto previamente apresentado.

A Parte IV apresenta a conclusão, trazendo considerações finais sobre a análise de equidade e comparando o impacto das duas formas de concessão.

Observa-se que, nesta tese, todas as referências bibliográficas são apresentadas apenas ao final, no Capítulo 12, para evitar eventuais repetições de referências, facilitar a consulta às mesmas e de modo a permitir melhor organização do trabalho.

1.4. Classificação desta pesquisa

Esta pesquisa se classifica como aplicada pela finalidade, quantitativa pela abordagem, exploratória pelos objetivos, *ex post facto* pela temporalidade e multidisciplinar pelos procedimentos. Além disso, é exploratória, quase-experimental e empírica².

Considera-se esta pesquisa aplicada, uma vez que ela apresenta metodologia passível de ser implementada, institucionalmente, permitindo balizar a administração pública em suas decisões e podendo alterar efetivamente a forma como se concede esses financiamentos.

É uma pesquisa quantitativa por se basear em métodos estatísticos para cálculo numérico do impacto, ou de métodos estatísticos para descrever desigualdades presentes nos investimentos públicos realizados, sem, contudo, buscar interpretação aprofundada daquilo que foi encontrado.

A pesquisa é exploratória, uma vez que apresenta método inovador, é quase-experimental, uma vez que usa de métodos artificiais de escolha entre os grupos de tratamento e de controle para inferir o impacto do tratamento, e é empírica, já que se busca demonstrar ser possível a aplicação de um método de cálculo para se chegar a um impacto da pesquisa.

Apresentada a classificação desta pesquisa, apresenta-se, a seguir, a hipótese e os objetivos das partes dois e três (análise da equidade e cálculo do impacto).

²<https://www.significados.com.br/tipos-de-pesquisa/#:~:text=Do%20ponto%20de%20vista%20da,%2C%20quantitativa%20ou%20quali%2Dquantitativa.>

1.5. Segunda Parte - Equidade

1.5.1. Introdução

A presença de minorias resulta em maior diversidade de ideias e estratégias para resolver diferentes tipos de problemas, pois, quando um grupo de pessoas tem que resolver problemas, os indivíduos trazem para a discussão suas diferentes informações, experiências e perspectivas únicas para resolver problemas. No caso particular das empresas, a diversidade está correlacionada com o aumento do valor da empresa e do lucro anual (Hunt et al. 2015, Dezso e Ross 2012, Dezso et al. 2017 APUD FERRARI et al., 2018). Quando um grupo de pessoas é exposto a um desafio que precisa de cooperação para ser resolvido, o melhor desempenho é obtido quando o grupo apresenta diversidade racial e de gênero. No caso particular de gênero, o movimento de ação afirmativa tem funcionado desde os anos 60 para aumentar a participação das mulheres em todos os campos. Infelizmente, o aumento dessa participação na ciência é lento (FERRARI et al., 2018).

Apesar dos diversos estudos que apontam o potencial ganho na ciência, tecnologia e inovação pela presença de mulheres, observam-se estudos que indicam um acesso limitado das mulheres na ciência. Uma busca no Site de Periódicos da CAPES encontrou 11.145 publicações com “mulheres” e “ciência”, sendo 8.039 em periódicos revisados por pares. De um modo geral, as pesquisas apontam para distintas formas de discriminação sexual contra as mulheres na ciência, incluindo uma menor presença das mulheres nas áreas de CTEM³ (Ciências, Tecnologias, Engenharia e Matemática) e o chamado “Teto de Vidro” (menor presença de mulheres em posições de liderança).

Apesar do número de estudos no assunto, encontrou-se indicador de discriminação sexual ainda pouco estudado ou explorado: mulheres recebem menos financiamento, em média e por projeto, do que homens. O encontrado está de acordo com diversos estudos que apontam que mulheres recebem menos que homens em seus empregos, mas inova ao apontar que isso também pode ser observado nas políticas públicas realizadas pelo CNPq. A misoginia permeia e se entranha até no setor público e se esconde sob a forma de uma discriminação pouco visível à sociedade. O menor financiamento leva a considerar uma forma de discriminação sexual que ainda não tinha sido observada em outros estudos e demonstra como essa discriminação pode

³O acrônimo mais usado é STEM, quem vem de: science, technology, engineering e mathematics (ciências, tecnologia, engenharia e matemática). Aqui, se usa o acrônimo em português por se tratar de documento que deve estar no idioma oficial brasileiro.

ser cultural, inconsciente e velada, uma vez que o valor financiado é pouco divulgado e uma vez que não foi observado que essa forma de discriminação é intencional.

Para confirmar a originalidade do estudo, foi realizada, no Portal de Periódicos da CAPES através de acesso da Rede Café, busca por “mulheres ciência cnpq”, “lattes impacto” e “mulheres lattes”. Todos os artigos encontrados foram analisados, não tendo sido encontrado estudo que aponte o fato acima narrado.

1.5.2. Hipótese

Ao buscar uma medida do impacto na Chamada Pública Universal MCTI/CNPq nº 14/2014, encontrou-se a seguinte proporção entre homens e mulheres:

Tabela 1 – Proporção entre Homens e Mulheres no Universal 2014

Sexo	Nº Solicitações		Demanda		Pagamentos Efetivados	
	Nº	% Total	Valor (milhões R\$)	% Total	Valor (milhões R\$)	% Total
F	5209	40%	481	38%	184	37%
M	7765	59%	794	62%	305	62%
s.d.	83	1%	8	1%	3	1%

Fonte: Autoria própria

Dessa forma verificou-se uma menor presença feminina que masculina. Mas, o que mais chamou atenção foi o fato de que essa presença era cada vez menor conforme a análise se aprofundava: nº solicitações -> valor demandado -> nº financiamentos -> valor demandado desses financiados -> valor efetivamente concedido. Conforme se aprofunda no nível medido, a participação feminina decresce cada vez mais.

Observou-se também que, quanto mais se aprofunda no nível medido, menor é a divulgação e transparência do dado em questão. Por exemplo, o número de bolsistas mulheres é amplamente divulgado, mas o valor financiado à elas não: divulga-se apenas o valor total. A constatação levantou a hipótese de que há misoginia presente no CNPq, misoginia essa disfarçada de modo que fica oculta a certas ações e que cresce na medida em que os dados são menos públicos/visíveis, presente principalmente num menor valor financiado às mulheres.

1.5.3. Objetivo

Dessa forma, o estudo de equidade objetivou verificar se há a discriminação sexual contra as mulheres, medindo o nível dessa discriminação e buscando indicar onde ela está presente.

1.5.4. Conclusão

A revisão bibliográfica encontra-se no capítulo 3 e o estudo se encontra no capítulo 4, no Artigo: Mulheres no CNPq – Uma análise de possíveis mecanismos de desigualdade de gênero no fomento federal à CT&I.

Encontrou-se a presença de discriminação sexual principalmente em ações de Apoio a Projetos de Pesquisa e em posições de liderança. Enquanto as mulheres são a maioria nas concessões realizadas (bolsas concedidas), são a minoria em termos de pesquisas científicas financiadas (como o Universal 2014). Além disso, elas recebem menos, no geral, que homens. Finalmente, encontrou-se que o simples fato de ser mulher implica, em média, num menor financiamento do projeto concedido.

1.6. Terceira Parte - Análise do Impacto

O uso da bibliometria permite, dentro da Comunicação Científica, avaliar a produtividade, suas áreas, recursos, processos e produtos, obtendo medidas a partir da análise estatística, apontando os impactos dos esforços e tornando a pesquisa visível e analisável. “Assim, os indicadores são instrumentos que, bem utilizados, dentro dos limites impostos pela técnica quantitativa de avaliação da ciência, podem funcionar como um filtro de controle e qualidade, dando acesso à estrutura e orientando a direção das políticas de gestão institucional” (SÔNIA DOMINGUES SANTOS BRAMBILLA; IDA REGINA CHITTÓ STUMPF, 2012). Nesse sentido, Currículo Lattes é uma ferramenta pública da qual se pode extrair mais de uma centena de indicadores por meio do uso de cientometria e ferramentas de ciência de dados e *Big Data*.

Tendo esses indicadores como base e sabendo-se o financiamento realizado, é possível buscar inferência causal e conseqüentemente uma medida do impacto do investimento público por meio de análise econométrica, obtendo medida de eficácia e custo-efetividade. Desenvolvido o método, esse pode ser aplicado para o cálculo do impacto do investimento em políticas públicas de financiamento à ciência. Como exemplos de políticas, temos a Chamada Pública Universal 2014 e a concessão de bolsas de doutorado, essas tanto pelo CNPq quanto pela CAPES.

1.6.1. Hipótese

É possível inferir causalidade entre indicadores retirados do Currículo Lattes, por meio de técnicas econométricas, de bibliometria e de ciência de dados em *Big Data*, tendo como uma das variáveis dependentes o financiamento público, de modo a medir o impacto desse financiamento por meio de um modelo econométrico quase-experimental com o uso do método das diferenças em diferenças. Dessa forma, é possível medir o impacto de uma política pública de investimento em ciência.

A análise realizada é escalável a várias políticas públicas, transparente, que permite comparação entre avaliações diferentes, tanto de forma longitudinal (entre políticas) quanto vertical (no tempo) e adequada a usos frequentes e automatizados ou semiautomatizados.

1.6.2. Objetivos

Desenvolver o modelo econométrico e aplicar a dois tipos de políticas públicas: concessão de bolsas de formação humana a nível de doutorado; e investimento direto em pesquisas científicas por meio de chamadas públicas, permitindo comparação entre essas políticas públicas.

1.6.3. Conclusão

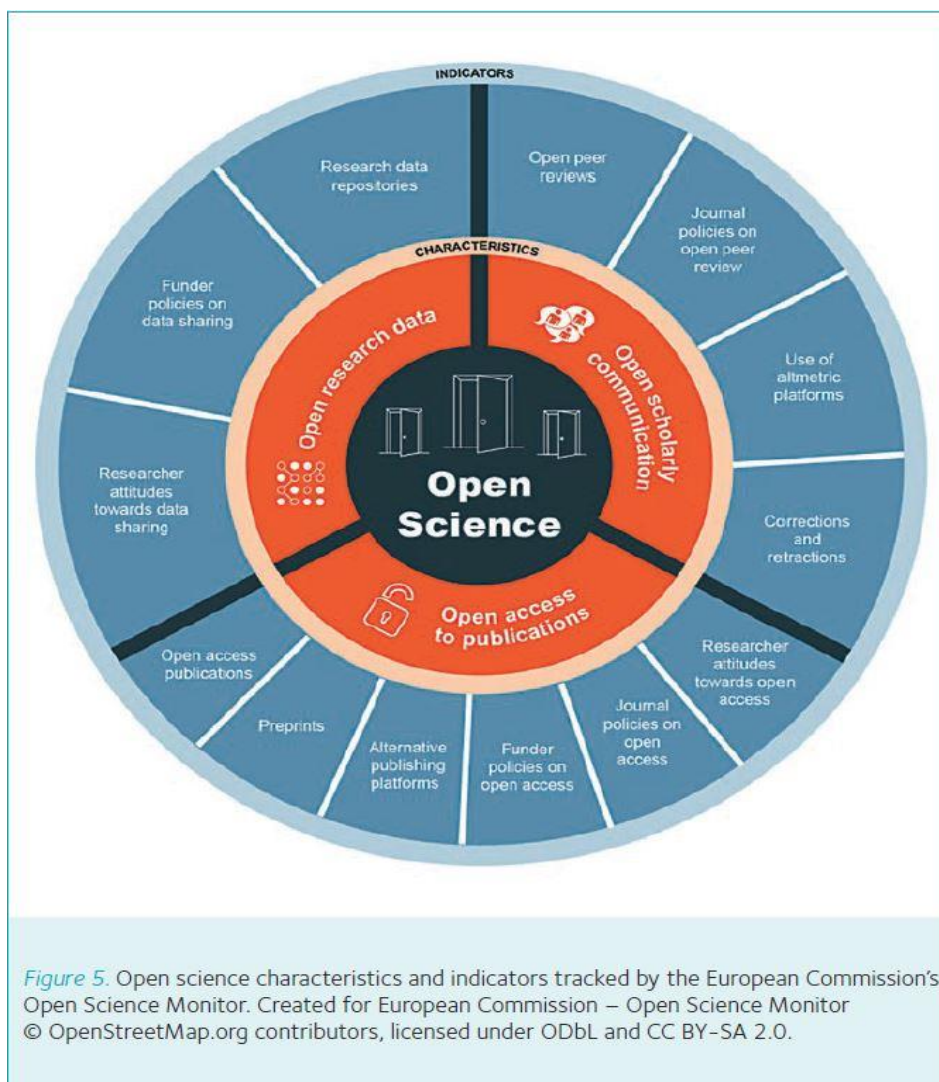
O modelo econométrico foi desenvolvido e utilizado, bem como o impacto calculado, tanto para a política pública de concessão de bolsas de doutorado como para investimento em pesquisas na Chamada Pública Universal MCTI/CNPq nº 14/2014.

O impacto medido foi expressivo, mas surpreende no baixo impacto em indicadores sociais, como novos livros, publicações em jornais e revistas, obras de arte, premiações etc. Ao final desta tese, os impactos são comparados e algumas conclusões tiradas.

2. ESTA PESQUISA NO CONTEXTO DA CIÊNCIA ABERTA

A Ciência Aberta é um movimento que incentiva a transparência da pesquisa científica desde a concepção da investigação até o uso de softwares abertos. Também promove esclarecimento na elaboração de metodologias e gestão de dados científicos, para que estes possam ser distribuídos, reutilizados e estar acessíveis a todos os níveis da sociedade, sem custos. Propõe, ainda, a colaboração de não cientistas na pesquisa, ampliando a participação social por meio de um conjunto de elementos que dispõem de novos recursos para a formalização da comunicação científica.

Figura 1 - A Ciência Aberta (Open Science) de acordo com a Comissão Europeia



Fonte: <https://www.physoc.org/magazine-articles/the-open-science-movement-revolution-is-underway/>

Esta pesquisa busca atender o modelo de ciência aberta, buscando publicar os artigos em revista de acesso aberto (*Open Access*), tornando os dados públicos (*Open Data*), permitindo sua reprodução (*Open Reproducible Research*) e tornando a programação realizada

pública (*Open Science Tools*). No entanto, encontrou-se limitações. Uma delas é que se trata de pesquisa em *Big Data* e, portanto, com impossibilidade de armazenamento gratuito pela maioria dos repositórios. Além disso, há dados personalizados, de modo que a LGPD impede a ampla publicação.

Dadas essas duas limitações, onde foi possível, os dados foram publicamente disponibilizados, permitindo o acesso aberto e a reprodução da pesquisa. Caso haja qualquer dado faltante, pode-se entrar em contato com este autor que suprirá os dados faltantes na medida do possível, mesmo que pela forma de link temporário em nuvem.

Com relação às ferramentas computacionais usadas, foram encontradas menções a ferramentas que analisam o Currículo Lattes automaticamente, como LattesDataXplorer (DE OLIVEIRA SANTIAGO; AFFONSO; MAGELA RODRIGUES DIAS, 2020), VantagePoint (RODRIGUES; GUIMARÃES, 2016), ScriptLattes⁴ e LucyLattes⁵.

No entanto, as ferramentas são de uso privado (VantagePoint), não puderam ser encontradas em formato de acesso aberto (LattesDataXplorer) ou não puderam ser aplicadas juntamente com o Extrator Lattes (LucyLattes). Dessa forma, tornou-se necessária a criação de uma ferramenta capaz de:

- Baixar os Currículo Lattes da ferramenta institucional Extrator Lattes;
- Descompactar o arquivo e tratar o XML gerado;
- Tirar do Currículo os Indicadores;
- Tirar do Currículo os Artigos Publicados;
- Armazenar as Informações em um Banco de Dados;
- Realizar as Regressões Lineares requeridas.

Para isso, foi construída ferramenta, de acesso aberto e licença GNU, disponível em https://github.com/AlbertoCamposSilva/CV_Lattes_Analyzer, que ainda está em produção. O código fonte é escrito em PostgreSQL e Python.

Lá também podem ser encontradas algumas das tabelas produzidas. Dessa forma, o estudo pode ser amplamente reproduzido e verificado, bem como a ferramenta utilizada para novos estudos. Qualquer outro dado necessário pode ser solicitado ao autor, que pretende continuar a desenvolver o sistema para permitir futuras análises.

Este estudo não foi submetido ao Comitê de Ética por se utilizar apenas de dados publicamente disponíveis (Currículo Lattes e informações do Portal de Dados Abertos do CNPq) ou anonimizados (Demanda Bruta e Atendida do Universal). Quanto a esses últimos,

⁴ <https://scriptlattes.sourceforge.net/>

⁵ <https://github.com/rafatieppo/lucylattes>

destaca-se de que, apesar de não estarem publicamente disponíveis, também são por força de lei dados públicos, de modo a reforçar a dispensa de análise pelo Comitê de Ética. Além disso, reforço que foram seguidos os preceitos éticos necessários para o desenvolvimento da pesquisa, entre eles a honestidade e precisão com relação aos dados coletados para evitar possíveis distorções de dados estatísticos que pudessem comprometer as interpretações. Nesse sentido, também é bom esclarecer que os dados publicamente disponibilizados estão anonimizados, a despeito de serem provenientes de dados publicamente acessíveis, por meio de substituição de qualquer id Lattes por um UUID único do qual este autor mantém tabela de relação.

2.1. Coleta de Dados

2.1.1. Currículo Lattes

Os Currículos Lattes foram baixados por meio da ferramenta institucional Extrator Lattes, disponível em <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>. A ferramenta é um webservice público com acesso restrito, de modo que apenas quem é cadastrado pode usar os serviços. Para o cadastro é necessário:

- Ser instituição de pesquisa;
- Fazer a solicitação;
- Ter IP fixo;
- Apresentar o projeto/a justificativa.

Maiores informações podem ser solicitadas pelo e-mail atendimento@cnpq.br. O serviço possui as seguintes funcionalidades:

- `getCurriculoCompactado`: Baixo o currículo compactado em formato XML cujas definições podem ser encontradas no seguinte link: <http://memoria.cnpq.br/documents/313759/1ee5aab5-84eb-4551-8dc2-c99764423c33>
- `getOcorrenciaCV`: retorna a ocorrência do último acesso, como, por exemplo:
 - Currículo recuperado com sucesso!
 - Nenhum currículo encontrado!
 - Wrong IP number. Can't get zipped Lattes from this IP number.
- `getDataAtualizacaoCV`: Retorna a última data de atualização do CVLattes;
- `getIdentificadorCNPq`: Retorna o ID Lattes baseado no CPF, nome e data de nascimento (necessários todos esses dados).

A lista de todos os IDs Lattes podem ser encontrados no seguinte link, provavelmente com atualização anual: <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados> .

2.1.2. Pagamentos Realizados

Os dados sobre todos os pagamentos realizados pelo órgão podem ser livremente baixados por meio do Portal de Dados Abertos do CNPq: https://dadosabertos.cnpq.br/pt_BR/dataset?tags=pagamento

As informações são fornecidas anualmente. Todas as tabelas relativas a pagamentos foram baixadas, tratadas e adicionadas a um banco de dados.

2.1.3. Demanda Bruta

Para uma das análises realizadas foi necessária tabela com a Demanda Bruta e Atendida do Edital Universal. As tabelas foram retiradas de uma página interna do CNPq, de acesso restrito apenas aos servidores do órgão: <http://relatorios.cnpq.br/jasperserver/login.html>

A ferramenta possui a funcionalidade de gerar dezenas de diferentes relatórios, entre eles, as demandas das chamada e informações sobre a pré-seleção (enquadramento técnico), se favorável (se tecnicamente meritória), a classificação e se foi concedida.

Embora os dados sejam de acesso restrito, é possível solicitar os mesmos por meio da Lei de Acesso à Informação (LAI), através do portal <https://www.gov.br/acessoainformacao/pt-br>. De qualquer forma, por não haver autorização expressa ou legal para a divulgação desses dados, esses não foram publicamente disponibilizados.

2.1.4. Sexo

A informação sobre o sexo é considerada como dado sensível pela LGPD e, portanto, o CNPq considera como dado não divulgável e não está publicamente acessível.

Por ser dado sensível, não é permitida a sua reprodução ou seu acesso aberto. Mesmo assim, poder-se-ia disponibilizar a mesma por meio de anonimização, transformando o ID Lattes (que permite a identificação do usuário) por um UUID (identificador único), o que foi feito onde a publicização não corria chance de ser revertida a uma possível identificação do pesquisador.

Indicamos que o método de retirada da informação sobre o sexo pode ser encontrado na função `Indicadores.getGender(names::list)`. A função pesquisa, além de no Banco de Dados construído, no site <https://api.genderize.io> que retorna o provável sexo da pessoa pelo primeiro

nome. Nota-se que há restrição de quantidade de acessos no site indicado, de modo que ele não pode ser usado para a procura do sexo a muitos nomes de uma única vez, embora tenha apresentado notável precisão quanto ao sexo encontrado.

2.1.5. Coleta de dados de financiamento por outras agências

As principais agências de fomento federais em CT&I são a CAPES e o CNPq. Enquanto uma agência pertence ao Ministério da Educação, a outra pertence ao Ministério da Ciência e Tecnologia. Talvez por isso, não há comunicação ativa entre as agências, troca de informações ou acessos a sistemas. A CAPES de fato envia uma planilha ao CNPq, mensalmente, com uma listagem de bolsas ativas para evitar duplo financiamento. No entanto, o acesso a essa listagem é restrito, mesmo para funcionários do CNPq, de modo que seu acesso não foi permitido. Assim, funcionários do CNPq não tem acesso a qualquer sistema da CAPES, informações de bolsistas ou pagamento realizados. Aliás, cita-se que foi tentado, por meio da Lei de Acesso à Informação (LAI) o acesso aos dados, que foi negado⁶.

No entanto, o pesquisador deve informar se recebeu bolsas em seu Currículo Lattes, pois tanto o CNPq quanto a CAPES especificam a obrigatoriedade em suas normas. Assim, a presença de eventual financiamento pela CAPES ou por outras agências de fomento pode ser retirado do Currículo Lattes, em campo específico para isso. Não há, contudo, informações sobre o montante financiado. Além disso, o Currículo Lattes é auto declaratório. Dessa forma, não se sabe quão precisos são os dados informados.

Ainda assim, por haver a obrigação normativa de inserção das informações no Lattes, espera-se razoável precisão. Assim, nota-se que, mesmo presentes alguns financiamentos não declarados na análise, esses não devem influenciar significativamente o resultado, pois espera-se um número reduzido desses, além de que, presentes, tendem a tornar o resultado subestimado e não sobre-estimado.

2.1.6. Período da análise

Registre-se que todos os estudos foram realizados utilizando-se os anos de 2014 e 2019, de modo a padronizar a análise, permitir comparação entre políticas públicas e evitar qualquer viés pela Pandemia COVID-19, que se iniciou em 2020. A exceção é o Artigo:

⁶ Este autor expressa aqui a repulsa pela dificuldade que se encontra em o pesquisador resgatar dados de agências públicas, mesmo a LAI dando expressamente a obrigação para isso e a LGPD permitindo, expressamente, a entrega dos dados, mesmo os sensíveis, para fins de pesquisas. Dessa forma, registra-se a repulsa pela recusa de órgãos públicos cumprirem o que este autor entende ser uma obrigação legal, além de moral.

Mulheres no CNPq – Uma análise de possíveis mecanismos de desigualdade de gênero no fomento federal à CT&I, em que algumas análises incluem esses anos.

Vale a pena mencionar que a Research Quality Framework do Development Advisory Group Australiano definiu como janela temporal para análise do impacto o período de seis anos (DONOVAN, 2008). Essa temporalidade está de acordo com este estudo, uma vez que os modelos econométricos comparam o início do ano de 2014 ao fim do ano de 2019, portanto, uma janela temporal de seis anos.

3. PARTE II - ANÁLISE DE EQUIDADE: MULHERES NA CIÊNCIA

3.1. Referencial Teórico

As mulheres são metade da população mundial e brasileira. Representam 74% dos Currículos Lattes cadastrados, mas ocupam apenas cerca de 1/3 das posições de liderança na academia científica. A discriminação sexual contra as mulheres na ciência é histórica e conhecida. Os números apontam para barreiras sociais contra a ocupação da mulher no corpo de elite em Ciência e Tecnologia. Rosa Alves et al. [s. d.] indicam que essas barreiras provavelmente se iniciam na primeira infância, fruto de tudo o que a criança é exposta pela mídia e faz parte da sociedade, reforçando o paradigma de que a ciência é feita por homens brancos. Em outro estudo, Rosa Alves et al. (2023) indicam que esse pré-conceito persiste até a graduação, em que muitos estudantes do sexo feminino indicam insegurança em entrar em cursos predominantemente masculinos (áreas CTEM), porque acreditam ser intelectualmente incapazes de acompanhar os pares masculinos, além de ter enfrentado situações desconfortáveis por causa do gênero feminino (comentários e assédio). De acordo com (BRITO et al., 2022), esse assédio é frequente na UFRGS (e provavelmente em outras universidades do país).

O encontrado por Rosa Alves *et al.*, em ambos os estudos, surpreende, pois, há a sensação de que a discriminação sexual está se reduzindo com os recentes ganhos da população feminina, sendo essa inclusive com formação média superior a masculina. Ainda assim, os estudos demonstram que a discriminação sexual se inicia na infância e persiste até a idade adulta, fortemente presente mesmo nos dias de hoje, fazendo com que as próprias mulheres acreditem ser intelectualmente inferiores aos homens.

O CNPq tem historicamente lutado contra essa discriminação, promovendo ações e políticas em favor da presença do corpo feminino. Como exemplo, um dos itens de avaliação para a escolha de membros dos Comitês de Assessoramento (CA) é a igualdade de sexo, raça e região, ordenamento que está presente no regimento do órgão⁷. Ainda assim, verifica-se alta desproporção entre homens e mulheres, tanto nos CAs quanto com bolsas PQ (DA SILVA et al., 2021) e que, ao longo do tempo, os grupos sub-representados mantiveram sua posição de sub-representatividade ao longo dos anos, seja o grupo feminino (na física, por exemplo) ou masculino (na enfermagem, por exemplo) (BEZERRA; STANISCUASKI; BARBOSA, 2022).

⁷ : "Art. 123. Para a composição dos Comitês de Assessoramento (CA) será considerada, sempre que possível, a representatividade das diversidades regional, institucional, de gênero, de cor, da área do conhecimento e de suas subáreas"

Hoje, 95% das professoras no primário são mulheres, no ensino secundário a porcentagem cai para 59% e na educação superior são apenas 47% (NODARI; SOARES DE OLIVEIRA; BARBOSA, 2023). Nota-se, portanto, como o número se reduz na medida em que o status aumenta e nota-se que essa porcentagem cai ainda mais para cargos de liderança. Num momento em que mulheres têm acesso à educação, a estrutura de desigualdade se mantém graças a estratégias socialmente construídas ao longo de séculos para perpetuar o domínio masculino.

Rodrigues et al (2016) indica que ainda são escassos os dados sistematizados sobre a formação, trajetória e perfil dos pesquisadores com foco no gênero, estrato social ou raça. Ainda, indicam que há formas sutis de discriminação que limitam a carreira da mulher, por meio de crenças e atitudes que operam inconscientes. Rodrigues et al (2016) utilizaram do Lattes para análise bibliométrica, medindo a relação entre os bolsistas PQ do CNPq e a presença das mulheres em cargos comissionados, encontrando que homens são mais produtivos que as mulheres, mas também a presença do “teto de vidro” na Fundação Oswaldo Cruz (Fiocruz). O teto de vidro “caracteriza-se pela menor velocidade na ascensão da carreira pelas mulheres, em comparação com a progressão profissional masculina, o que resulta na sub-representação das mulheres nos postos de tomada de decisão e, conseqüentemente, limita o alcance de posições de maior prestígio e o acesso a uma melhor remuneração” (RODRIGUES; GUIMARÃES, 2016).

Em relação ao Teto de Vidro, Borges et al. (2019) indica forte sub-representação feminina em empresas do setor de energia e ainda que:

- A participação das mulheres no mercado de trabalho é menor do que a dos homens;
- Os salários das mulheres são ainda inferiores aos dos homens (77,5% do salário médio dos homens), apesar de possuírem, em média, mais habilitações (formação acadêmica) do que os homens, segundo o IBGE (Agência Brasil, 2018);
- Sub-representatividade das mulheres em posições de tomada de decisão, especialmente nas áreas de alta tecnologia;
- Baixa representação na política brasileira. No Congresso Nacional, por exemplo, apenas 15% do total de parlamentares são mulheres. Com um número insuficiente de mulheres nessas posições, projetos e temas-chave para o empoderamento feminino e diversidade são comumente negligenciados;
- Violência contra a mulher, incluindo feminicídio e assédio sexual.

(BORGES DE LIMA FACÇÃO et al., 2019)

Um estudo mais recente apontou que mulheres em posição de liderança publicam aproximadamente a mesma quantidade de artigos por ano e apresentam quase o mesmo tempo

decorrido na carreira, enquanto as mulheres produzem mais recursos humanos. Ainda assim, ocupam apenas cerca de 1/3 dos cargos de pró-reitora de pesquisa e que, “para que as mulheres cheguem ao topo, é necessário que se adaptem à mesma métrica destinada aos homens, enfrentando desafios adicionais que indicam um teto de vidro persistente” (NODARI; SOARES DE OLIVEIRA; BARBOSA, 2023). Ou seja, de um modo geral, os estudos apontam que as barreiras que buscam impedir o acesso das mulheres nas posições de liderança e poder são invisíveis, socialmente construídas e possivelmente inconscientes.

Outros termos existem para as diversas formas sociais de ceifar a presença feminina na ciência: *leaky pipeline* (vazamento no fluxo), *firewall*, segregações horizontais e verticais. *Leaky pipeline* descreve o gradual desaparecimento da presença feminina na medida de progressão da carreira e *firewall* “se refere-se às formas sistêmicas e dinâmicas de discriminação que compõem a própria estrutura das organizações e estão por toda parte”. Segregação horizontal descreve a maior presença masculina em determinadas áreas, normalmente mais valorizadas, enquanto a vertical se correlaciona ao Teto de Vidro e à *Leaky pipeline*. Ainda há o sexismo cultural usado por Savigny (2014), que descreve uma barreira invisível e significativa que impede a progressão da mulher na academia (NODARI; SOARES DE OLIVEIRA; BARBOSA, 2023) e o termo labirinto de cristal (LIMA, 2007) representando “os constantes obstáculos invisíveis que as mulheres enfrentam para conseguirem avançar nestas carreiras. O labirinto de cristal é uma alegoria para um trajeto repleto de desafios para as mulheres, mas transparente para o restante da sociedade” (ALVES; BARBOSA; LINDNER, 2023).

A presença feminina não encontra obstáculos apenas na ciência. Um estudo apontou baixa presença feminina na produção de patentes (PERUCCHI; MUELLER, 2014), e outro que, muito embora não haja evidências de distinção entre homens e mulheres na proporção no país de formação acadêmica (no país ou exterior), há evidências de que homens publicam mais artigos no exterior (ROCHA-CARPIUC; MADEIRA, 2020).

No contexto paulista do agronegócio, há forte diferenciação de produção acadêmica nas gerações mais envelhecidas (≥ 55 anos) e paridade nas gerações mais jovens, além das mulheres possuírem maior incremento médio de publicação anual, apontando para tendência de paridade ou mesmo reversão em favor das mulheres na produtividade científica. Ou seja, o estudo sugere que pode estar havendo declínio da misoginia. No entanto, o mesmo estudo observou que, quanto mais alto é o cargo, menor é a penetração feminina, sendo que os argumentos da qualificação e produtividade são insuficientes para explicar o fenômeno (OLIVEIRA; SERRA, 2018). Apesar dessa tendência observada de paridade de gênero no

número de publicações entre a população mais jovem, outros estudos não apontam para tendência de igualdade em posições de poder ou *status* nas próximas décadas, que novamente faz inferência a existência de um labirinto de cristal.

No sentido de demonstrar como o teto de vidro é um construto social, invisível e de difícil explicação, Nodari *et al.* (2023) indicou que, para chegar em posições de prestígio, como a de decano em universidades, mulheres não precisam ter uma maior produtividade do que homens. Isso significa que os requisitos entre os distintos gêneros são equivalentes, muito embora mulheres passem por maior dificuldade em se atingir esses requisitos. Assim, chegam na conclusão de que igualdade de requisitos para atingir a posição de prestígio não significa igualdade de gênero, uma vez que há barreiras sociais que dificultam que as mulheres atinjam esses requisitos (NODARI; SOARES DE OLIVEIRA; BARBOSA, 2023).

À primeira vista, pode parecer haver igualdade de oportunidades e de requisitos e que as mulheres não alcançam a posição de liderança por serem menos produtivas ou menos capazes. Dessa forma, para entender melhor o motivo por trás de tantas indicações de misoginia, mesmo podendo a mulher alcançar a mesma posição se tiver a mesma produtividade, foi procedida entrevistas com três pesquisadoras, que indicaram que:

- Mulheres são menos chamadas para colaborações científicas, mesmo quando mais qualificadas para a pesquisa em questão que outros colaboradores do sexo masculino;
- Mesmo quando estão fisicamente mais próximas, ou sendo virtualmente mais acessíveis, de um modo geral os pesquisadores buscam parceiros do sexo masculino para suas colaborações, mesmo que sejam estrangeiros e menos acessíveis;
- Como há forte presença masculina, a mulher precisa utilizar da linguagem e do assunto de interesse masculino para chamar atenção à sua pesquisa, enfrentando um labirinto de cristal. Ou seja, como a pesquisa é um construto social, os grupos sociais automaticamente e inconscientemente buscam favorecer a pesquisa que utiliza da mesma linguagem social. Esse favorecimento desfavorece as mulheres mesmo quando a análise é realizada às cegas, por exemplo;
- Mulheres ainda são as principais cuidadoras da família, de modo a ter carga de horário de trabalho muito maior que a masculina;
- Há a tendência na escolha masculina a cargos de poder sempre que há paridade de currículo, ou mesmo quando o currículo feminino apresenta melhores

indicadores que o masculino, mas esses são indicadores “menos valorizados” (como orientações de graduação);

- Mulheres enfrentam “brincadeiras” (assédio) constantes, que as desestimulam a permanecer no meio social masculino e, portanto, a progressão na carreira ou assunção de cargos de poder.

O encontrado acima é resultado de entrevistas informais e não pode ser considerado um estudo, e, portanto, nem dito como verdadeiro. Mas, essas explicações dadas por três experientes pesquisadoras em distintas áreas do conhecimento surpreenderam e permitiram uma maior compreensão de como a misoginia funciona, de modo invisível, indetectável à sociedade, mas ainda ceifando a presença feminina e por isso foi aqui reproduzida. O narrado condiz com os estudos apresentados e demonstram como as mulheres enfrentam barreiras invisíveis para chegar a um patamar de produtividade semelhante ao masculino. Um estudo ainda será realizado, formalizando essas descobertas.

O sexismo cultural pode levar a um ciclo vicioso. Mulheres são impedidas de atingir prestígio, de modo que ficam menos visíveis à sociedade. Com menor visibilidade, há maior dificuldade com a produtividade científica e menor inserção da mulher na ciência. Assim, as mulheres possuem um menor incentivo a prosseguir na carreira científica e há, com isso, redução da inserção da mulher na ciência. O ciclo pode levar a perdas de talentos, menor produtividade e eventual perda econômica. A igualdade de gênero poderia aumentar em cerca de US\$ 12 trilhões o PIB global até o ano de 2025, de acordo com um relatório do McKinsey Global Institute (WOETZEL et al., 2015 apud NODARI; SOARES DE OLIVEIRA; BARBOSA, 2023).

Assim, os estudos mostram que as desigualdades de oportunidade entre homens e mulheres permeiam a sociedade em diferentes esferas: a política, a econômica e a social. Demonstram que há discriminação sexual, socialmente construído, muitas vezes invisível e permeado na maneira que os indivíduos leem o mundo, perdurando por séculos, persistente e longe de ser resolvido (ARÊAS; SANTANA; BARBOSA, 2020). Nesse sentido,

atividades de educação e de divulgação científica, muitas vezes, levam à reprodução em vez de mitigar as desigualdades sociais. Falar de equidade neste contexto diz respeito a compreender que há diferenças e desigualdades que estruturam o campo da ciência e da divulgação científica que precisam ser levadas em conta para garantir que todas e todos tenham a mesma dignidade e exercício da cidadania em meio a uma cultura científica. Falar em equidade de gênero é ir além do enfoque quantitativo da necessidade de ter mais mulheres na ciência, mas criticar em vias de transformar as práticas dominantes excludentes dentro dos campos em questão. Requer, portanto,

desafiar as estruturas de poder que criam desigualdades e inibem a participação e inclusão nessas práticas.

(...)

Compreender os múltiplos fatores que afetam a participação das mulheres nas ciências pode ajudar a construir caminhos para uma educação que busque dar conta dessas iniquidades históricas.

(ANA PAULA ALVES PEREIRA et al., 2022, p. 23–24)

3.2. Metodologia

Foi realizada análise dos investimentos do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) para homens e mulheres. Dessa forma, os investimentos realizados foram analisados ao todo e depois agrupados por ano, por faixa de valor, por área de conhecimento e por linha de fomento, sendo essa última ainda subdividida em Apoio a Projetos de Pesquisa (APQ) e Bolsas Produtividade (PQ).

A pesquisa demonstrou a presença do teto de vidro, confirmando os diversos estudos em que as mulheres não conseguem atingir posições de liderança. Mas, adiciona, indicando que o teto de vidro também está presente nos projetos de pesquisa de alto valor, nas áreas de conhecimento ditas como prioritárias (CTEM) e nas linhas de fomento.

Também se constatou a presença do labirinto de cristal, uma vez que mulheres precisam ter o mesmo resultado que homens, mesmo recebendo um valor a menor eles, já considerada a média por projeto. O encontrado está de acordo com os estudos que indicam que as mulheres recebem menos que os homens em postos de trabalho, mas inova ao demonstrar o mesmo fato em financiamentos públicos a projetos de pesquisa.

3.2.1. Delta – Uma medida de Elasticidade

Num financiamento público para pesquisa (como por exemplo, no edital UNIVERSAL 2014 lançado pelo CNPq), concorrem homens e mulheres. Como mulheres são a maioria da população brasileira, e como hoje são a maioria no ensino superior, também deveriam ser a maioria nessa concorrência. Isso nem sempre é verdade, pois a misoginia se intrinca dentro da cultura social, fazendo com que as próprias mulheres se depreciem, acreditando não serem capazes de concorrer. Por exemplo, no UNIVERSAL 2014, concorreram 15.125 mulheres e 18.773 homens (com 216 sem sexo definido), 44,6% de mulheres apenas.

Se cerca de 45% das mulheres concorrem a uma chamada pública de financiamento à pesquisa, cerca de 45% dos contemplados deveriam ser de mulheres. No entanto, ao analisar a

proporção dos beneficiados, encontrou-se 5.209 mulheres que ganharam a concorrência e 7.765 homens (com 83 de sexo indefinido). Ou seja, entre aqueles que foram financiados, tem-se apenas 40,1% de presença feminina.

A misoginia poderia indicar que há diversos fatores por trás da observação realizada, incluindo uma possível menor produtividade feminina ou uma possível inferioridade curricular feminina em comparação à masculina. Ou, talvez, os projetos apresentados fossem inferiores. Assim, a misoginia cultural declararia que, como a chamada pública busca a excelência, essa se encontra mais concentrada no corpo masculino e, por isso, são os homens os mais contemplados. Pura falácia. Quando analisamos os pagamentos realizados, apenas dos contemplados, temos a seguinte proporção entre homens e mulheres:

Tabela 2 – Proporção entre Homens e Mulheres no Universal 2014

Sexo	Número de Pesquisadores		Demanda		Pagamentos Efetivados	
	Nº	% Total	Valor (milhões R\$)	% Total	Valor (milhões R\$)	% Total
F	5209	40%	481	38%	184	37%
M	7765	59%	794	62%	305	62%
s.d.	83	1%	8	1%	3	1%

Fonte: Autoria própria

Observa-se que, no Universal 2014 e embora mulheres sejam 51% da população brasileira, mulheres representam apenas 45% da concorrência, 40% dos contemplados, que elas solicitaram apenas 38% do total solicitado, mas que foram contempladas com apenas 37% dos recursos do edital. Ao analisar a curva decrescente torna-se clara a misoginia em toda a cultura científica: não só são homens mais presentes, desde o início do processo, mas a presença feminina diminui a cada passo dado, sendo que a transparência também diminui em cada passado.

A constatação surpreendeu, uma vez que há a política contra a misoginia presente no CNPq. Como exemplos dessa política, temos que o regimento interno do CNPq demanda: "Art. 123. Para a composição dos Comitês de Assessoramento (CA) será considerada, sempre que possível, a representatividade das diversidades regional, institucional, de gênero, de cor, da área do conhecimento e de suas subáreas". Observa-se que os CAs são o corpo científico que escolhem quais propostas possuem mérito e realizam a classificação delas nas chamadas públicas de Apoio à Pesquisa do CNPq. Dessa forma, a norma busca a equidade num corpo fundamentalmente diretivo do órgão.

Outros exemplos são os editais que pontuam favoravelmente a presença feminina (como, por exemplo, a Chamada CNPq/MCTI/FNDCT Nº 23/2022 - implementação do Programa de PD em Nióbio - INOVANIÓBIO), ou os editais lançados especificamente para

promover as mulheres, como a Chamada CNPq/MCTIC N° 31/2018 - Meninas nas Ciências Exatas, Engenharias e Computação. Finalmente, há a cultura interna que busca orientar e garantir a equidade: os CAs são orientados a tomar cuidado com isso e relatórios são muitas vezes realizados ao final da ação incluindo análise de equidade.

No entanto, observa-se que essas ações que diretamente promovem a inserção da mulher na academia científica são ainda poucas, esparsas e insipientes. Muito embora a equidade esteja prevista no regimento interno, há forte presença masculina na composição dos CAs. As chamadas destinadas ao público feminino são poucas e não ocorrem de forma contínua, anualmente. Finalmente, a orientação interna de busca de equidade fica fragilizada ante a forte cultura de se buscar a “excelência” da pesquisa.

O CNPq vê a busca da excelência como a busca por investir no melhor projeto possível, bem como naquele pesquisador que possui as melhores chances de “sucesso” na pesquisa. Dessa forma, é o Currículo Lattes a principal ferramenta de análise para essa medida de excelência, verificando-se nele o histórico do pesquisador e os indicadores presentes no currículo, para a escolha “do melhor” a ser financiado. Ocorre que, pelo levantamento bibliográfico acima levantado, essa forma de agir desfavorece a equidade. Assim, embora o órgão busque promover a equidade, tanto por meio informativo, quanto por meio normativo, nota-se ainda forte desigualdade, tanto regional quanto sexual.

Para um estudo quantitativo dessa desigualdade, é útil o cálculo do quanto cada gênero recebe, em média por projeto financiado. Essa é uma medida de elasticidade, definida como “[o percentual de alteração em uma determinada variável, dada uma variação percentual em outra. Pode ser relacionado com sensibilidade ou reação da variável em questão em relação a outras](#)”⁸.

No exemplo do Universal 2014 acima apontado, 40% das contemplações são de mulheres e essas recebem apenas 37% dos recursos concedidos. Assim:

Equação 1 – Elasticidade dos Recursos Femininos no Universal 2014

$$Elasticidade = \frac{\% \text{ recursos}}{\% \text{ mulheres}} = \frac{37\%}{40\%} = 92,5\%$$

Ou seja, em média, mulheres recebem apenas 92,5% dos recursos que os homens recebem, já corrigido para a proporção entre homens e mulheres financiadas. É como se, pelo simples fato de ser mulher, o financiamento concedido caia em cerca de 8%; ou o financiamento aumente em cerca de 8% pelo simples fato de se ser homem. Comparar 37% com 40% pode

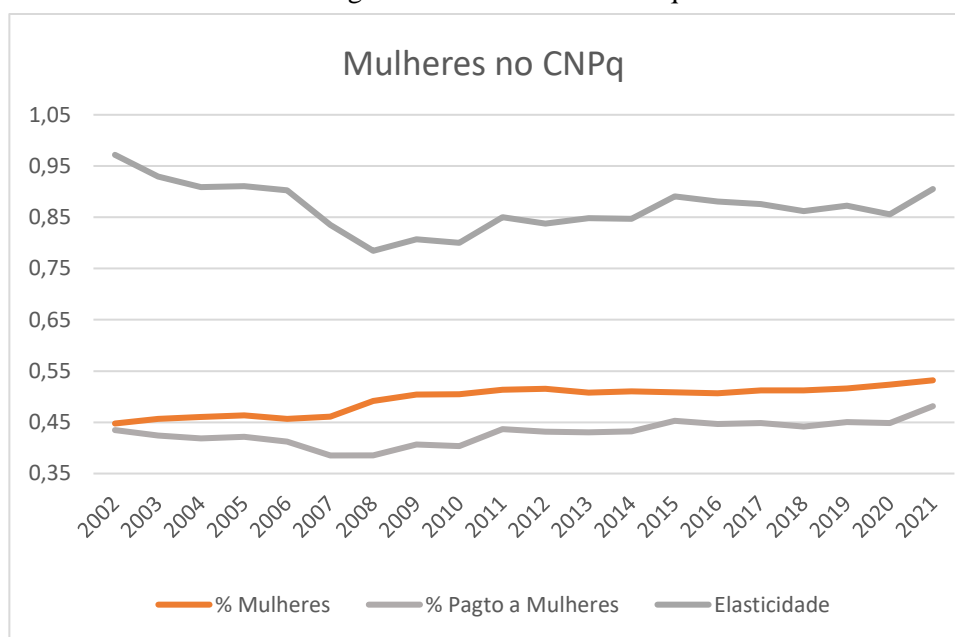
8

<https://www.sun0.com.br/artigos/elasticidade/#:~:text=O%20que%20C3%A9%20elasticidade%3F,quest%C3%A3o%20em%20rela%C3%A7%C3%A3o%20a%20outras>.

levar a uma falsa impressão de que há “apenas” 3% de diferença entre os sexos, o que não é verdade.

Como as mulheres são 40% dos financiados, as mulheres então recebem 92,5% * 40% dos recursos, ou seja, 37% desses. O conceito de elasticidade é muito útil no caso, pois consegue representar bem uma dupla desvantagem da mulher: são apenas 40% das financiadas e, além disso, mesmo que fossem exatamente 50%, ainda receberiam apenas 92,5% dos recursos.

Figura 2 – Mulheres no CNPq



Fonte: Autoria Própria.

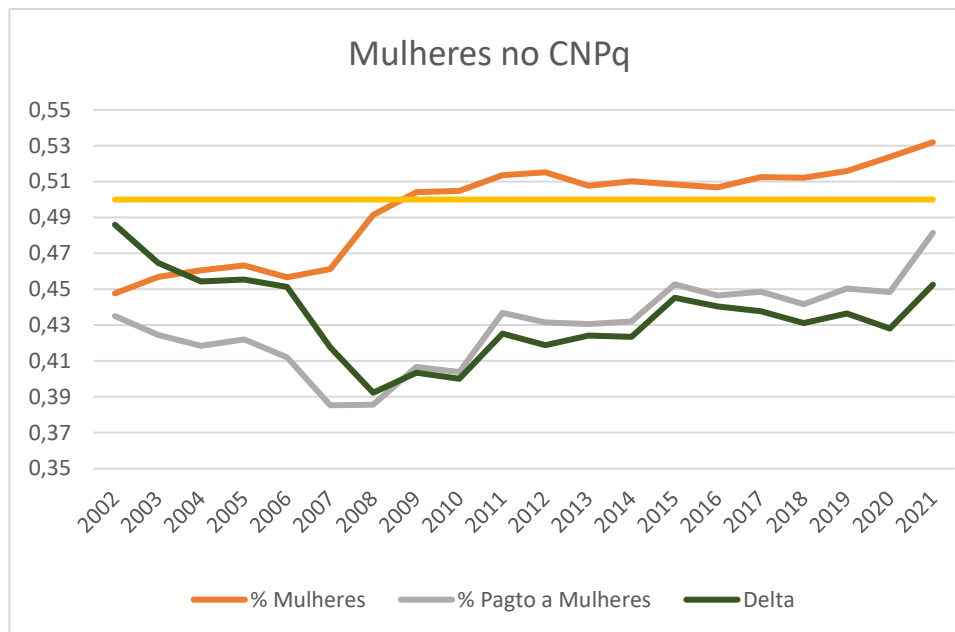
Ocorre que a equidade na elasticidade é atingida quando essa é igual a um (o que se denomina de unitária). Mas, a equidade é atingida quando temos 50% dos financiamentos para homens e 50% dos financiamentos a mulheres, ou seja, a equidade dos indicadores adotados é atingida quando esses são iguais a 0,5. O fato de que a equidade na elasticidade é atingida quando essa se iguala a 1, mas a equidade é atingida nos outros indicadores quando esses se igualam a 0,5, dificulta a interpretação do gráfico acima: temos que olhar cada indicador de um modo diferente.

A solução é simples, basta aplicar uma transformada na elasticidade, dividindo-a por 2. Dessa forma, será calculado o Delta, dividindo-se a elasticidade acima apresentada por 2 (Equação 3). Este conceito de Delta será repetido em diversas análises, motivo pelo qual é aqui esmiuçado.

Equação 2 - Elasticidade

$$Delta = \frac{\% \text{ do valor recebido por mulheres}}{2 * \% \text{ Mulheres beneficiadas}} = \frac{37\%}{80\%} \cong 0,46$$

Figura 3 - Mulheres no CNPq



Fonte: Autoria Própria.

Observe como a interpretação do gráfico acima é mais simples. Podemos traçar a linha em 0,5, de modo a facilmente verificar onde há ou não desvantagem para as mulheres. Podemos, com a transformada, verificar que o número de concessões favorece as mulheres a partir de 2009, mas que essas têm, historicamente, recebido menos recursos que os homens.

A aplicação de Delta neste estudo deriva de um estudo de Corak (2020) em que propõe uma medida de elasticidade para comparar a mobilidade da equidade entre gerações. Delta é, portanto, uma medida de elasticidade, se constituindo de uma divisão de porcentagens (GREENLAW et al., 2017, seção 5.1). O fator de 2 a mais no denominador, presente apenas para facilitar a visualização nos gráficos que serão mostrados, não altera a natureza do indicador e permite que ele seja apresentado graficamente em conjunto com os outros indicadores.

Neste exemplo, Delta pode ser interpretado como, numa escala de 0 a 100, a fatia de recursos que cada um fica se já considerada a proporção financiada entre homens e mulheres. Como exemplo, acima o valor de Delta é de 46%, ou seja, supondo que o mesmo número de mulheres e homens sejam financiados, mulheres ficam com uma fatia de 46% dos recursos em média por projeto (valor de Delta) enquanto homens ficam com 54% dessa média, uma diferença total de 8%.

A medida de Delta é importante, pois são três fatores correlacionados entre si, mas com significados diferentes. Suponha, por exemplo, que mulheres recebam, em determinada área, 70% dos financiamentos, mas que Delta vale apenas 0,33. Isso significaria que, apesar de

receberem mais projetos de pesquisa, recebem em cada projeto apenas 1/3 do que receberiam se fossem do sexo masculino, mesmo recebendo quase 50% dos recursos dispendido (o que poderia ser erroneamente interpretado como equidade).

Assim, quanto maior for a distância de Delta de seu eixo (de 0,5 ou 50%), maior é a tendência à desigualdade. Nota-se, portanto, que um Delta de 0,40 entre o valor pago às mulheres e o número de mulheres que obteve financiamento é um possível indicador da presença de padrões de desigualdade de gênero (misoginia).

4. ARTIGO: MULHERES NO CNPQ – UMA ANÁLISE DE POSSÍVEIS MECANISMOS DE DESIGUALDADE DE GÊNERO NO FOMENTO FEDERAL À CT&I

Artigo Aceito pelo XIV ENPEC:



AVALIAÇÃO DO TRABALHO

Prezado/a **ALBERTO DE CAMPOS E SILVA**,

Informamos que seu trabalho intitulado "MULHERES NO CNPQ – UMA ANÁLISE DE POSSÍVEIS MECANISMOS DE DESIGUALDADE DE GÊNERO NO FOMENTO FEDERAL À CT&I", foi avaliado e considerado **ACEITO SEM RESSALVAS** para apresentação no XIV ENPEC. O trabalho recebeu o parecer a seguir:

Parecer do Avaliador 1

Interessante e pertinente, com boa análise dos dados coletados.

Parecer do Avaliador 2

A temática e a discussão que o trabalho propõe é urgente e fundamental para a sociedade e o meio acadêmico. O estudo tem um robusto corpo de dados e excelente análise. Trabalho aceito sem ressalvas

Parecer final:

O trabalho foi aceito para apresentação no XIV ENPEC.

Linha Temática (LT): LT 11 - Políticas Educacionais e Currículo

Autor/es: ALBERTO DE CAMPOS E SILVA e ALEXANDRE GUILHERME MOTTA SARMENTO

Informamos ainda que a versão final do trabalho, com os nomes dos autores, será a mesma enviada na submissão e não poderá ser substituída.

Atenciosamente,
Comissão Organizadora do XIV ENPEC

Mulheres no CNPq – Uma análise de possíveis mecanismos de desigualdade de gênero no fomento federal à CT&I

Women at CNPq – An analysis of possible mechanisms of gender inequality in the federal funding of ST&I

Alberto de Campos e Silva

Universidade Federal do Rio Grande do Sul (UFRGS)

silva@albertocampos.com.br

Alexandre Guilherme Motta Sarmiento

Universidade Federal do Rio Grande do Sul (UFRGS)

agmotta.ufrgs@gmail.com

4.1. Resumo

Este trabalho detectou a presença de padrões de desigualdade de gênero ao estudar o financiamento realizado pela principal agência de fomento à ciência no Brasil, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Verificou-se que os números absolutos de financiamentos às mulheres aumentaram ao longo do tempo, mas há indícios da presença de padrões de desigualdade de gênero na forma de subfinanciamento do valor em projetos de pesquisa, principalmente naqueles de maior monta, e na forma de menor presença em cargos de liderança. Percebe-se que tais padrões podem ser perpetuados pela desigualdade na distribuição de gêneros nos maiores responsáveis por avaliar os projetos (Bolsistas Produtividade).

Palavras chave: discriminação sexual, igualdade de gênero, mulheres na ciência, avaliação de políticas públicas, mulheres na CTEM, CNPQ.

Abstract

This work detected the presence of patterns of gender inequality by studying the funding carried out by the main agency for the promotion of science in Brazil, the National Council for Scientific and Technological Development (CNPq). It was found that the sheer

numbers of funding to women increased over time, but there are indications of the presence of patterns of gender inequality in the form of money underfunding in research projects, mainly in those of larger scale, and by lesser presence in leadership positions. It is noticed that such patterns can be perpetuated by the inequality in the distribution of genders in those most responsible to evaluate the projects.

Key words: Sexual Discrimination, Gender Equality, Women in Science, Evaluation of Public Policies, Women in STEM, CNPq.

4.2. Introdução

Uma das principais ferramentas de melhoria social a longo prazo está na política pública de investimento em educação. Psacharopoulos & Patrinos (2018) estimam o retorno econômico desse investimento em mais de 8% ao ano, além de encontrar que o retorno da educação feminina é maior que o masculino, que esse é ainda maior em países em desenvolvimento e que está provavelmente subestimado. No entanto, padrões de desigualdade de gênero continuam a apresentar obstáculos a que os educandos atinjam todo o seu potencial.

Se considerarmos apenas um único fator, gênero permanece aquele que é o mais importante para determinar as escolhas na escola, emprego, indústria de inserção, salário, oportunidades de treinamento recebidas e aproveitadas, promoções, pensão e assim por diante (LI et al., 2022). De acordo com as Nações Unidas, mulheres representam 33,3% de todos os pesquisadores, 12% como membros de academias nacionais de ciência, 22% em áreas de ponta como inteligência artificial, tendem a ter carreiras menores, com salário menor, com menor representação e recebem menos promoções (International Day of Women and Girls in Science, United Nations, [s. d.]). Nesse mesmo sentido, Coe et al. (2019) indicam a existência de uma cultura penetrante, persistente e prejudicial de assédio que limita a participação e o avanço de mulheres em ciência, tecnologia, engenharia e matemática (CTEM) e indicam que deve haver uma mudança na cultura e no clima organizacional para combater o fato.

Se existentes, tais padrões de desigualdade de gênero seriam uma afronta ao princípio de igualdade da Constituição Federal Brasileira de 1988 (BRASIL, 1988). A Constituição, já em seu primeiro artigo, traz a dignidade da pessoa humana como fundamento e em seu 3º como objetivo fundamental reduzir as desigualdades sociais, sem preconceitos de sexo, entre outros. Finalmente, o artigo 5º, considerado cláusula pétrea, fundamenta a igualdade de gênero:

Art. 5º Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a

inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade, nos termos seguintes:

I -homens e mulheres são iguais em direitos e obrigações, nos termos desta Constituição; (...)

XLI -a lei punirá qualquer discriminação atentatória dos direitos e liberdades fundamentais; (...)

(BRASIL, 1988)

Para Coe et al. (2019), a única forma de progredir na resolução do problema cultural de discriminação sexual de forma permanente é uma mudança cultural. Para que essa ocorra, o primeiro passo é evidenciar a existência da cultura discriminatória, para depois promover formas voluntárias e normativas de mudanças culturais (COE et al., 2019). A análise que se faz é a de equidade, ou seja, uma análise sobre como os grupos minoritários, sociais são afetados, quais grupos arcam com os custos (SILVA, 2015).

O CNPq é o órgão brasileiro responsável por “fomentar a pesquisa científica, tecnológica e de inovação e promover a formação de recursos humanos qualificados para a pesquisa, em todas as áreas do conhecimento” (CNPQ, 2020). Dessa forma, dos órgãos de abrangência federal, é o CNPq aquele que melhor pode evidenciar a presença de padrões de desigualdade de gênero no financiamento à pesquisa.

Esta pesquisa analisou a possível presença de padrões de desigualdade de gênero, por meio de uma análise de equidade no CNPq. Após esta introdução e breve introdução ao CNPq, especifica-se como os dados foram coletados e são introduzidos os primeiros dados sobre a presença da mulher no CNPq. São estudados os financiamentos por faixa de valores e áreas do conhecimento, para depois promover o estudo financiamento a projetos de pesquisa (APQ) e de Bolsas Produtividade (bolsas PQ), uma vez que esses últimos são os principais líderes na escolha dos financiamentos. Finalmente, apresentamos uma conclusão.

4.3. O CNPq

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) tem “por finalidade promover e estimular o desenvolvimento da investigação científica e tecnológica em qualquer domínio do conhecimento” e “é pessoa jurídica subordinada direta e imediatamente ao Presidente da República, terá sede na Capital Federal e gozará de autonomia técnico-científica, administrativa e financeira, nos termos da presente lei” (Lei nº 1.310). Criado em 1951, é uma das instituições mais sólidas na área de investigação científica e tecnológica entre os países em desenvolvimento. O CNPq é vinculado ao Ministério da Ciência e Tecnologia (MCTI), mas, enquanto o primeiro fomenta a ciência no país com financiamento a INCTs, com

a concessão de bolsas e com o financiamento a projetos de pesquisa, o segundo instituiu a política de CT&I no Brasil. O orçamento do CNPq gira em torno de pouco mais de um bilhão de reais anuais, representando 0,04% dos gastos públicos.

Apesar do orçamento relativamente pequeno se comparado aos totais dos gastos públicos, o CNPq representa um importante papel na ciência brasileira. O CNPq atua principalmente concedendo duas modalidades de financiamento, Bolsas e Auxílios, embora seu orçamento se componha também do custo administrativo do órgão, isenções fiscais de importação de produtos destinados à CT&I, entre outros menores.

4.4. Coleta dos dados

Os dados desta análise de equidade foram obtidos por meio do portal de dados abertos do CNPq, publicamente disponível em <http://dadosabertos.cnpq.br>. As planilhas apresentadas apresentam os seguintes dados: Ano Referência, Processo, Beneficiário, Linha de Fomento, Modalidade, Categoria/Nível, Nome Chamada, Programa CNPq, Grande Área, Área, Subárea, Instituição Origem, Sigla UF Origem, País Origem, Instituição Destino, Sigla Instituição Destino, Sigla Instituição Macro, Cidade Destino, Sigla UF Destino, País Destino, Valor Pago.

As tabelas dos anos de 2020 e 2021 foram solicitadas por este autor e enviadas pelo órgão. As planilhas foram baixadas e importadas em Banco de Dados relacional Postgree SQL, o que permitiu que diversas tabelas derivadas fossem montadas por meio do uso da linguagem de programação SQL. Esses foram importados para uma planilha eletrônica e os gráficos gerados.

Observa-se, no entanto, que nas planilhas importadas não há informações sobre o Sexo ou a Raça. Apesar disso, as planilhas contêm o nome completo do beneficiário, de forma que o sexo pode ser deduzido a partir desse. Para tal, foi providenciada tabela com o nome completo de mais de sete milhões de pessoas, com informações do sexo incluídas pelo próprio usuário. Essa planilha foi retirada do próprio Lattes, de modo que não se espera erro. Um código SQL foi gerado para comparar os nomes das duas planilhas, anexando a informação sobre o sexo nas planilhas do portal do CNPq.

Ao longo deste texto, “Mulheres” se referem a pessoas que se declararam do sexo Feminino, “Pessoas” ao total de pessoas sem distinção de sexo e “Homens” a pessoas que se declararam do sexo “Masculino”. Dessa forma, este estudo não busca trazer questões de gênero, aqui tratado como sinônimo de sexo, uma vez que usualmente aquilo que se declara é o sexo do nascimento.

4.4.1. Delta

Para estudo da questão, é útil o cálculo do quanto cada gênero recebe, em média por projeto financiado. Para isso, será calculado o Delta, dividindo-se a porcentagem do valor recebido pelas mulheres com duas vezes a porcentagem de mulheres beneficiadas (Equação 3). Este conceito de Delta será repetido em diversas análises, motivo pelo qual é aqui esmiuçado.

Equação 3 - Elasticidade

$$\Delta = \frac{\% \text{ do valor recebido por mulheres}}{2 * \% \text{ Mulheres beneficiadas}} = \frac{43\%}{108\%} \cong 0,40$$

A aplicação de Delta neste estudo deriva de um estudo de Corak (2020) em que propõe uma medida de elasticidade para comparar a mobilidade da equidade entre gerações. Elasticidade é um conceito de economia que mede a capacidade de resposta de uma variável a mudanças em outra variável. Delta é, portanto, uma medida de elasticidade, se constituindo de uma divisão de porcentagens (GREENLAW et al., 2017, seção 5.1). Neste estudo, estamos usando um fator de 2 a mais no denominador, presente apenas para facilitar a visualização nos gráficos que serão mostrados. Assim, uma elasticidade menor que 0,5 indica uma relação inelástica, elástica se maior que 0,5 e unitária se igual a 0,5.

Neste exemplo, Delta pode ser interpretado como, numa escala de 0 a 100, a fatia de recursos que cada um fica se já considerada a proporção financiada entre homens e mulheres. Como exemplo, acima o valor de Delta é de 40%, ou seja, mulheres ficam com uma fatia de 40% dos recursos em média por projeto (valor de Delta) enquanto homens ficam com 60% dessa média. Isso significa que mulheres, em média, receberam por projeto apenas 80% dos recursos que homens receberam (o dobro de Delta), ou que homens recebem 125% daquilo que recebem as mulheres ($0,8^{-1}$).

Assim, quanto maior for a distância de Delta de seu eixo (de 0,5 ou 50%), maior é a tendência à desigualdade. Nota-se, portanto, que um Delta de 0,40 entre o valor pago às mulheres e o número de mulheres que obteve financiamento é um possível indicador da presença de padrões de desigualdade de gênero, pois o financiamento concedido às mulheres é, em média, apenas 80% daquele concedido aos homens.

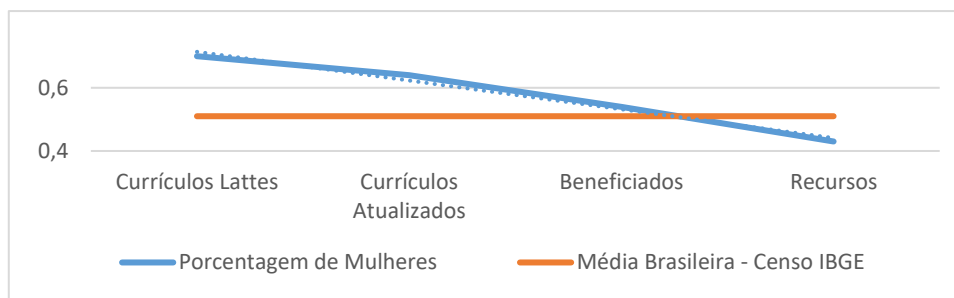
4.5. A mulher no CNPq

Ao longo dos anos, o CNPq vem adotando série de medidas com o intuito de combater qualquer cultura discriminatória contra as mulheres. Entre as medidas encontradas, estão a

promoção de chamadas exclusivas à participação de mulheres (Meninas na Ciência, Futuras Cientistas), prorrogação de bolsas por motivo de parto e/ou adoção, inclusão no Lattes da Licença Maternidade e incentivo a critérios de julgamento que favoreçam a mulher em casos de parto e/ou adoção. Além disso, o CNPq tem lançado critérios e chamadas que favorecem a presença de mulheres (apenas como exemplo, a Chamada nº 23/2022, InovaNióbio, pontua a presença de mulheres na equipe de pesquisa).

De acordo com o Censo Brasileiro realizado pelo IBGE, o Brasil tem predominância de mulheres, sendo 51% da população brasileira. No início de 2022, o Currículo Lattes tinha 7.427.277 registros, sendo que as mulheres representavam 70% dos cadastros realizados. Se considerarmos apenas os currículos atualizados a partir de 2020, temos 1.630.326 mulheres e 916.105 homens cadastrados, de modo que as mulheres continuam predominando com 64% dos cadastros. Se verificarmos a proporção de mulheres e homens beneficiados por quaisquer ações do CNPq, temos uma proporção de 54% de mulheres. No entanto, ao verificarmos o valor concedido a ambos os gêneros, verificamos que as mulheres recebem apenas 43% dos recursos do órgão.

Figura 4 - Porcentagem de Mulheres na Ciência



Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

4.5.1. O Financiamento às Mulheres ao Longo dos Anos

Para acompanhar a tendência da existência, aumento ou diminuição de padrões de desigualdade, separamos a porcentagem da participação das mulheres em termos de beneficiados das ações e participação dos recursos captados, obtendo a Figura 5.

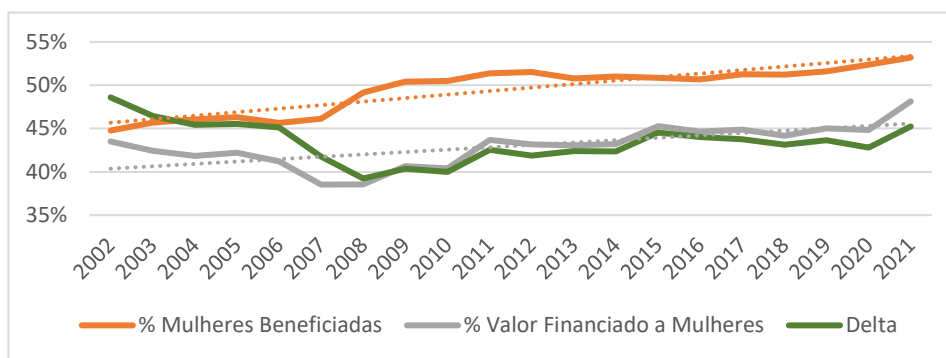
Na figura, o menor valor de Delta é de 39% em 2007, ou seja, mulheres ficam com uma fatia de 39% dos recursos em média por projeto (valor de Delta) enquanto homens ficam com 61% dessa média (mulheres, em média, receberam por projeto apenas 78% dos recursos que homens receberam).

Quanto à porcentagem de mulheres beneficiadas e do valor financiado às mulheres, para a manutenção da equidade cada linha do gráfico acima deveria se aproximar do valor de

51% (senso IBGE). Observa-se que o número de mulheres beneficiadas aumentou de 2002 a 2021, passando de 45% a 53% e, portanto, indicando forte tendência à eliminação de qualquer padrão de desigualdade de gênero.

A análise do valor financiado, ou de Delta, indica o contrário. Eliminando-se o ano de 2021 da análise, para eliminar possíveis efeitos da pandemia COVID-19, temos que apenas 45% dos recursos foram destinados às mulheres, sendo que Delta foi de apenas 0,44 em 2019. O valor de Delta indica que mesmo em 2019 as mulheres receberam, em média, 14% a menos que os homens em cada financiamento realizado. Acompanhando o gráfico nota-se que a partir de 2008, Delta se aproximou cada vez mais de seu eixo de 0,5, mas que essa aproximação se encerrou em 2015, mantendo-se quase linear a partir dessa data ou mesmo uma tendência ao aumento da distância de Delta ao eixo.

Figura 5 - Participação das Mulheres nas Ações do CNPq ao longo do tempo



Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

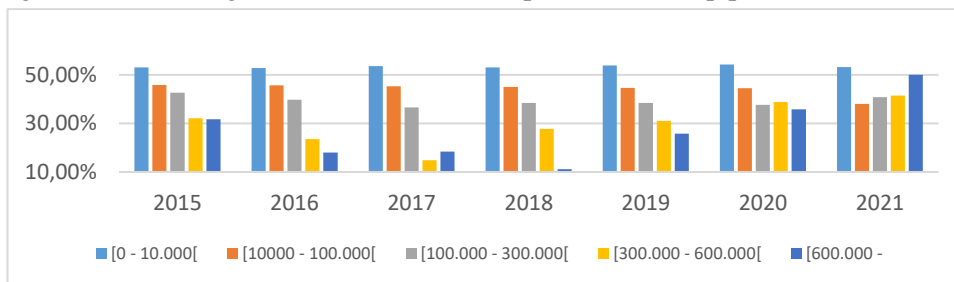
Portanto, pode-se concluir que, apesar de haver redução em qualquer padrão de desigualdade de gênero em relação ao número de concessões às mulheres, há padrão de desigualdade de gênero em relação ao montante pago em média por projeto e que esse padrão está, desde 2015, sem qualquer tendência à melhoria (descartado o ano de 2021, por visualmente representar ponto de inflexão, ser *outlier* e possível presença de viés pela Pandemia COVID-19).

4.5.2. Mulheres no CNPq por Faixa de Valores

Para analisarmos mais profundamente qualquer mecanismo de discriminação, precisamos ainda analisar se isso ocorre com concessões de alto valor, ou de baixo valor. Dessa forma, agrupamos os valores concedidos por faixas e calculamos a porcentagem de recursos concedidos às mulheres. As faixas foram escolhidas para representar de modo aproximado a variação com que o CNPq financia suas ações e buscou ser distribuída conforme os diversos

financiamentos realizados, como, por exemplo, na faixa [0 – 10.000[que se constitui, principalmente, de pequenos financiamentos como bolsas de Iniciação Científica a graduandos (bolsas IC). Dessa forma, realizando a análise, temos o seguinte gráfico:

Figura 6 – Porcentagem de Mulheres contempladas no CNPq, por Faixa de Valores

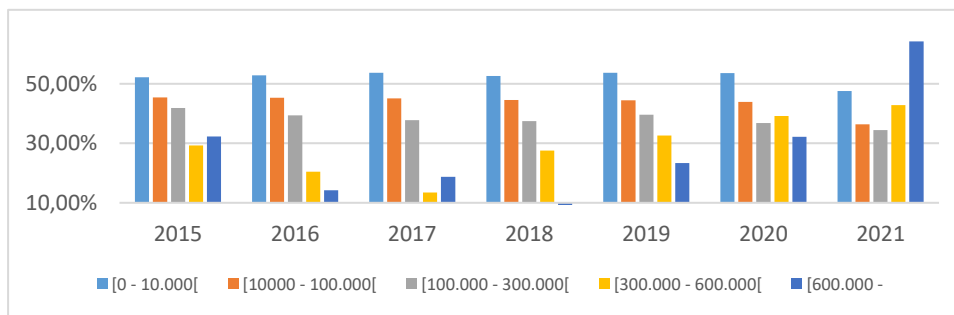


Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

Podemos observar que, exceto na primeira faixa, as mulheres foram menos contempladas com financiamentos de forma sistemática e contínua no tempo. Observa-se também que, em geral, embora mais mulheres que homens sejam beneficiadas ao total, essas receberam um maior número de financiamentos apenas na primeira faixa de valores. Ainda, observamos tendência a que essa diferença seja ainda mais acentuada conforme o valor das faixas aumenta.

Se fizermos análise semelhante para o valor financiado às mulheres, obtemos o seguinte gráfico:

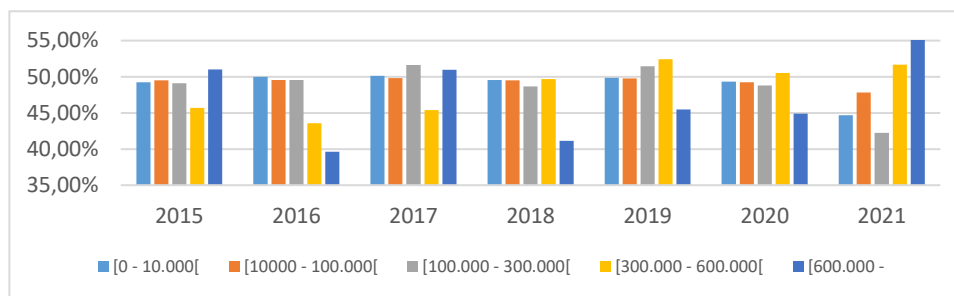
Figura 7 - Porcentagem de Recursos Concedidos à Mulheres no CNPq, por faixa de valores



Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

Novamente, exceto pela primeira faixa, a mulher recebeu menos recursos que homens, em todas as faixas de valores (descartado o ano de 2021). O gráfico acompanha, de certa forma, o anterior (Figura 6), tornando necessário analisar o Delta:

Figura 8 - Delta, por faixa de valores



Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

Retirando-se da análise a última faixa de valor do ano de 2021, por se tratar de um ponto atípico e ocorrido durante a Pandemia COVID-19, observa-se Delta muito próximo ao eixo de 0,5 (unitário), com apenas leve tendência a estar abaixo dessa linha. A exceção se dá, normalmente, em projetos de alto valor (R\$ 300.000,00 ou mais). Dessa forma, observa-se que, para pequenos projetos e bolsas, não há aparentemente mecanismos de desigualdade, mas que esses se concentram em grandes projetos, fazendo com que mulheres recebam, em média, por projeto, menos recursos que homens.

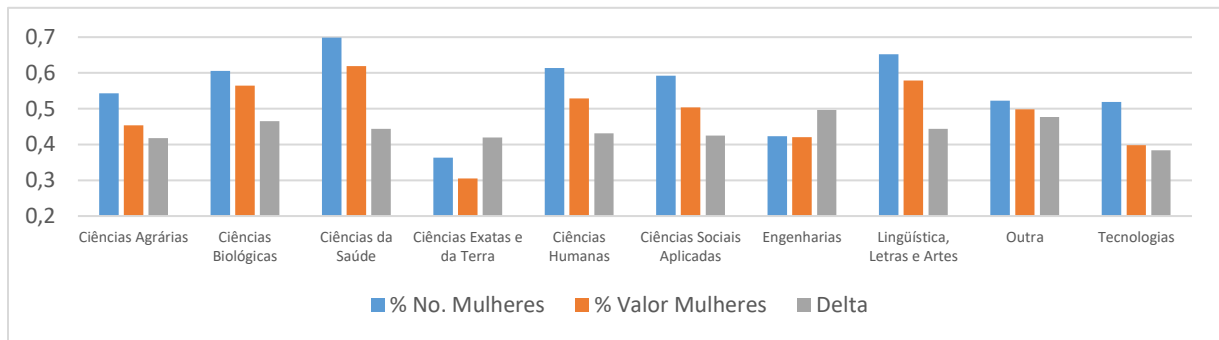
Assim, nota-se duplo efeito: mulheres são mais contempladas com financiamentos pequenos e menos com grandes. Além disso, quando recebem financiamentos grandes (R\$ 300.000,00 ou mais), tendem a receber uma fatia menor desses do que homens.

4.5.3. Mulheres no CNPq por Área do Conhecimento

No CNPq as áreas do conhecimento são divididas em Grande Áreas, Áreas e Sub-áreas. As Grandes Áreas são: Ciências Agrárias, Ciências Biológicas, Ciências da Saúde, Ciências Exatas e da Terra, Ciências Humanas, Ciências Sociais Aplicadas, Engenharias, Linguística, Letras e Artes e Tecnologias, sendo que cada Grande Área possui suas Áreas e Sub-Áreas. Dessa forma, é possível realizar o estudo da presença de padrões de desigualdade de gênero em cada Grande Área, ao realizar a comparação por essas áreas.

Como já visto, a área de ciência, tecnologia, engenharia e matemática (CTEM) requer especial atenção, já que é historicamente reconhecida por ser discriminatória contra as mulheres. Por outro lado, há indícios de que o inverso ocorra para as áreas humanas e da saúde (GARCÍA-PEÑALVO et al., 2022). Dessa forma, fizemos a comparação do número de concessões às mulheres, em cada Grande Área do Conhecimento:

Figura 9 - Mulheres no CNPq por Área - 2021

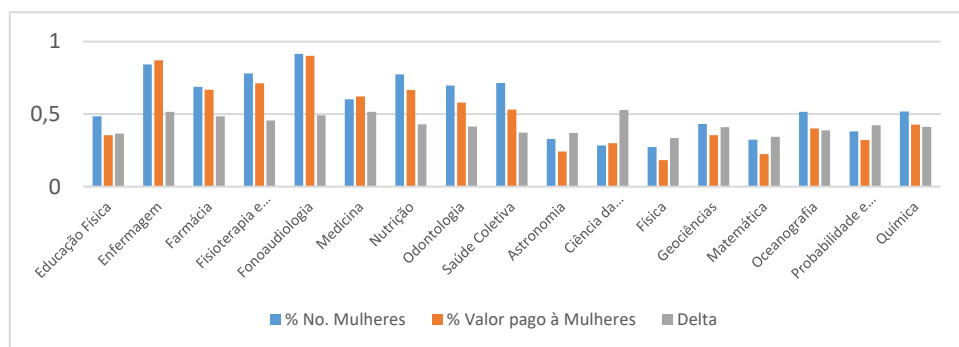


Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

Primeiramente podemos observar, pelo gráfico acima, que delta é inelástico (menor que 0,5) em todas as áreas, exceto a de Engenharias, cuja participação de mulheres é de apenas 42%, de forma que há aparente desfavorecimento às mulheres em todas as áreas. Ainda, percebe-se dois extremos: Ciências da Saúde, com 70% de participação das mulheres, e Ciências Exatas e da Terra, com 36% de participação delas.

Para uma melhor compreensão, separamos apenas essas duas Grande Áreas e plotamos os dados para cada Área:

Figura 10 - Participação das Mulheres - Saúde e Exatas - 2021



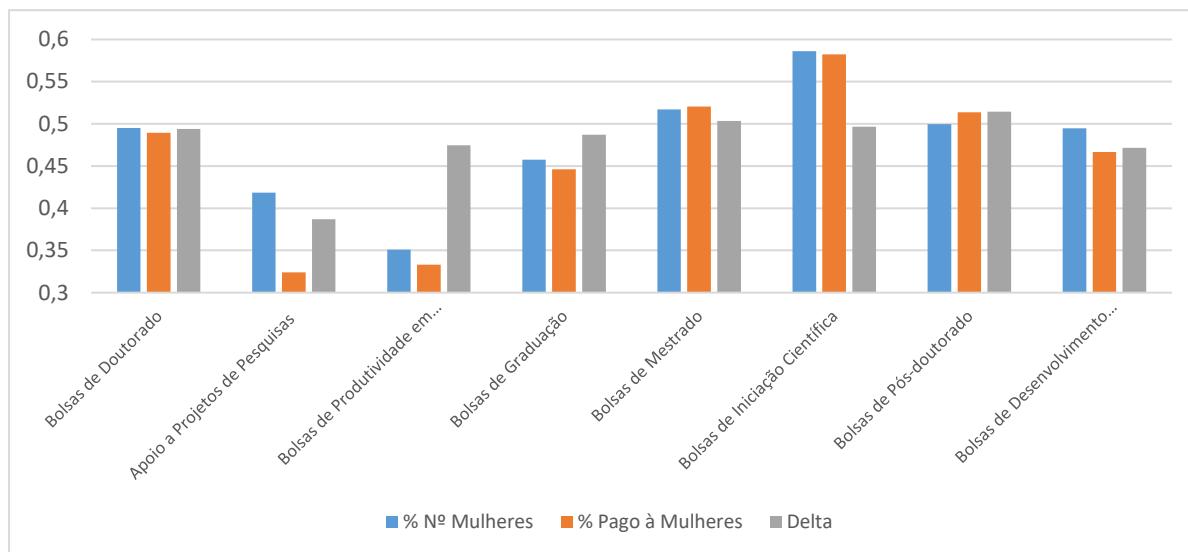
Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

Assim, nas áreas da saúde, podemos observar predominância feminina, tanto em número de concessões, quanto em valor recebido, exceto para a área de Educação Física. O exato inverso ocorre para as áreas das Ciências Exatas e da Terra, em que há clara predominância masculina, exceto para Oceanografia e Química, onde há certa equidade. Além disso, observamos Delta menor que 0,5 em todas as áreas, tanto da Saúde quanto de Exatas e da Terra, exceto para Medicina e Ciência da Computação. No entanto, em Ciência da Computação temos forte presença masculina.

4.5.4. Presença das Mulheres por Linha de Fomento

Linhas de fomento são divisões internas ao CNPq que indicam formas distintas de financiamento. Temos, como exemplos de linhas, Bolsas de Graduação, Mestrado e Doutorado, Apoio a Projetos de Pesquisas (APQ), entre outras. Dessa forma, pode ser útil a análise das linhas de fomento para verificar se há mecanismos concentrados em uma ou outra forma de concessão.

Figura 11 - Participação das Mulheres no CNPq, por Linha de Fomento - 2002 a 2021

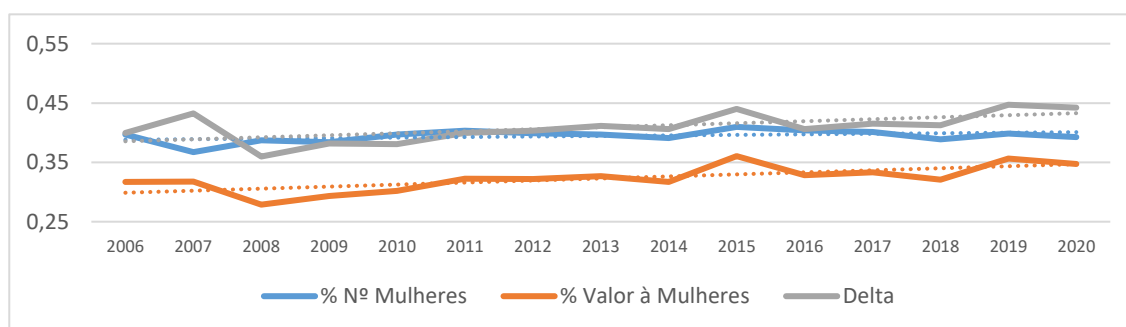


Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

A análise indica que Bolsas de Doutorado, Mestrado e de Pós-Doutorado estão, em regra, bem distribuídas. Há leve predominância masculina em Bolsas de Desenvolvimento Tecnológico (DTI, último grupo no gráfico) e em Bolsas de Graduação. No entanto, o desequilíbrio no Apoio a Projetos de Pesquisas (APQ) e Bolsas de Produtividade em Pesquisa (PQ) é nítido e forte. Assim, cabe análise pormenorizada dessas duas modalidades.

4.5.5. Apoio a Projetos de Pesquisas (APQ)

Figura 12 - Participação das Mulheres em APQ no CNPq, por ano.



Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)

APQ representa o financiamento a projetos de pesquisa e, além de possivelmente bolsas, compreende as rubricas de Custeio e Capital. Apenas a título de exemplo, a chamada com o maior valor financiado da linha APQ foi o “Edital MCT/CNPq 02/2006 – Universal”, no valor total pago de R\$ 127.835.754,00. Dessa forma, APQ representa uma das principais ferramentas para o cumprimento da missão institucional do CNPq e é a segunda linha de pesquisa que mais recebeu financiamento, sendo responsável por 19% dos gastos do órgão.

Nota-se que a participação de mulheres em APQ no CNPq é praticamente constante com o tempo e aproximadamente igual a 40% nos últimos anos. Apesar de levemente ascendente, a porcentagem dos recursos distribuídos às mulheres subiu de 30% a 35%, apenas 5% ao longo de 14 anos, de modo que Delta nunca passou dos 45%.

4.5.6. Bolsas Produtividade (PQ)

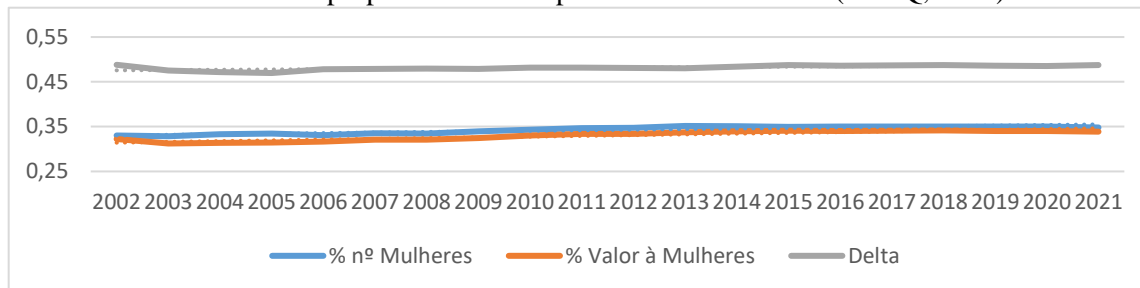
A Bolsa Produtividade (PQ) é “Destinada aos pesquisadores que se destaquem entre seus pares, valorizando sua produção científica segundo critérios normativos, estabelecidos pelo CNPq, e específicos, pelos Comitês de Assessoramento (CAs) do CNPq” (Portal CNPq - RN-028/2015). Nesse contexto, a bolsa PQ do CNPq torna-se um processo de reconhecimento de seus pares e, portanto, um “signo de inter-reconhecimento no campo, que situa determinados agentes no topo da hierarquia acadêmica, podendo ser considerado um importante elemento de distinção” (OLIVEIRA et al., 2022).

Bolsistas PQ são considerados, portanto, como uma elite entre os pesquisadores brasileiros e os bolsistas de nível 1 seriam a elite dessa elite. Portanto, a presença de padrões de desigualdade de gênero entre os bolsistas PQ contribui para a manutenção dessa cultura em todo o CNPq:

Cabe ressaltar que os agentes que acumulam mais capital acumulado em determinado campo [bolsistas de produtividade nível 1A] também são aqueles com maior capacidade de “deformar” a ordem do campo, operando de forma ativa na determinação das “regras do jogo”, tendencialmente a partir de posições ortodoxas que visam a manutenção de suas posições de poder. No caso dos bolsistas PQ, essa posição fica ainda mais evidente considerando que apenas aqueles de nível 1, que são indicados pela comunidade acadêmica é que podem compor os comitês de assessoramento, possuindo, assim, uma capacidade mais efetiva de alterar as regras do jogo (OLIVEIRA et al., 2022, páginas 192 e 193).

No CNPq, temos que apenas 35% das bolsas PQ concedidas são para às mulheres sem tendência ao longo do tempo de melhoria gradual. Um estudo mais detalhado revela que a proporção de bolsas PQ de nível 1 é ainda menos equânime (OLIVEIRA et al., 2021).

Figura 13 - CNPq - Mulheres e Bolsas Produtividade por ano
 Fonte: autoria própria baseada no portal de dados abertos (CNPQ, 2022)



4.6. Conclusão

O estudo permitiu comparar o financiamento realizado pelo CNPq de vários pontos de vistas diferentes. O crescente número de concessões às mulheres, que chegou a ultrapassar a linha dos 51%, indica uma distribuição das concessões mais igualitária por parte do CNPq. Percebe-se presença feminina equânime (ou mesmo predominante) na base (processos de baixo valor e bolsas em geral, exceto a PQ).

No entanto, o crescente aumento ao número de mulheres financiadas não é acompanhado pelo respectivo aumento no valor total financiado por projeto: há tendência que as mulheres recebam menos recursos que homens, principalmente onde liderança é exigida (projetos de alto valor, APQ e Bolsas Produtividade). Assim, como disse Coe et al. (2019), há indicativo de uma cultura penetrante, persistente e prejudicial de assédio que limita a participação e o avanço de mulheres. No entanto, encontrou-se também que essa cultura não se limita às áreas de CTEM, uma vez que mulheres recebem menos recursos (em média por projeto) que homens, em quase todas as áreas.

Coe et al. (2019) reforça que deve haver uma mudança na cultura e no clima organizacional para combater a desigualdade. Já que os comitês de assessoramento formam o principal corpo consultivo do CNPq e já que esses são formados principalmente por Bolsistas Produtividade, esses são principalmente aqueles responsáveis pela escolha de quais projetos de pesquisa serão ou não aprovados e pela escolha do valor a ser financiado. Assim, a presença de padrões de desigualdade de gênero em Bolsas Produtividade tende a levar à perpetuação desses mesmos padrões no CNPq como um todo.

Dessa forma, o primeiro passo para a redução de mecanismos discriminatórios seria corrigir a equidade no principal corpo consultivo do órgão (Bolsistas Produtividade), que por si só poderá corrigir outro ponto de forte divergência, os financiamentos a projetos de pesquisa (APQ), principalmente nas faixas de valores mais altas. Para Coe et al. (2019), uma liderança inclusiva é necessária para a mudança institucional. Essa mudança na liderança pode ser obtida

através de instrumentos voluntários ou legislativos. Entre os instrumentos voluntários teríamos as metas voluntárias e o ativismo político. Dos três métodos propostos, cota foi o menos preferido e metas voluntárias, bem como pressão pública e ativismo político, foram considerados mais eficientes. Apesar disso, caso os instrumentos falhem, cotas podem ser aplicadas e estudos na África do Sul indicam ser um método eficiente na promoção da igualdade de gênero (COE et al., 2019).

Finalmente, percebemos equidade de distribuição entre homens e mulheres, em termos de número de financiamentos realizados (ou mesmo favorecimento às mulheres). Dessa forma, qualquer atitude direcionada à área CTEM, na forma de incentivar a maior participação feminina, deve vir acompanhada de política de incentivo de participação masculina nas áreas de saúde, biológicas, sociais e letras.

5. Parte III - Análise do Impacto da Política Pública de investimento na Ciência

5.1. Definição do Impacto da Ciência

Um primeiro passo para o desenvolvimento de um novo método para a análise do impacto da ciência é o estudo de como esse impacto é avaliado por agências de fomento ao redor do mundo. O impacto da pesquisa também pode ser definido como:

Quão bem a atividade contribui ao avanço do conhecimento enquanto estimula o ensino, a formação e a aprendizagem? Até que ponto a atividade proposta amplia a participação das minorias (por exemplo, sexo, etnia, deficiência, geográfica, etc.)? Até que ponto isso vai aumentar a infraestrutura para pesquisa e educação, como instalações, instrumentação, redes e parcerias? Os resultados serão amplamente divulgados para melhorar a compreensão científica e tecnológica? Quais podem ser os benefícios da atividade proposta para a sociedade?

(HOLBROOK; FRODEMAN, 2010, p. 218 apud BORNMANN, 2013b, p. 224, tradução nossa).

De acordo com Donovan (2008), a ciência pode trazer benefícios sociais, culturais, econômicos ou ambientais. O benefício Social representa a melhora na qualidade de vida, incluindo a de grupos minoritários (equidade), em todas as esferas, incluindo, por exemplo, as esferas social, política, da ciência e da cultura, saúde e segurança. O benefício Econômico é a melhora da economia social, o crescimento econômico, incluindo o acréscimo de capital humano, social, real e cultural. O benefício Ambiental pode ser definido como qualquer melhora ao meio ambiente e ao estilo de vida, incluindo a preservação ambiental, redução de descartes e da poluição, preservação ou mesmo acréscimo da biodiversidade e ecossistemas, preservação do clima ou mesmo redução de malefício ambiental, como redução no consumo de combustíveis fósseis. O benefício Cultural é o acréscimo cultural, como entendimento de nossas origens, do funcionamento de nossa sociedade, nossas relações internas e externas, preservação histórica e cultural.

O uso da bibliometria permite, dentro da Comunicação Científica, avaliar a dimensão da ciência produzida, suas áreas, recursos, processos e produtos, obtendo medidas a partir da análise estatística, apontando os impactos dos esforços e tornando a pesquisa visível e analisável. “Assim, os indicadores são instrumentos que, bem utilizados, dentro dos limites impostos pela técnica quantitativa de avaliação da ciência, podem funcionar como um filtro de controle e qualidade, dando acesso à estrutura e orientando a direção das políticas de gestão institucional” (SÔNIA DOMINGUES SANTOS BRAMBILLA; IDA REGINA CHITTÓ STUMPF, 2012)

Dito isso, a meta inicial é analisar como o mundo usa da cientometria para realizar possível análise do impacto da ciência.

5.2. A Análise do Impacto da Ciência no Mundo

Com base nessas definições de impacto da ciência, houve, por parte das agências governamentais ao redor do mundo, diversas tentativas de mudar a forma como se avalia a ciência, principalmente com a intenção da escolha de como fazer o investimento público de forma a ter o maior retorno social o possível. Dessa forma, passa-se a uma pequena análise de como o impacto da pesquisa é avaliado ao redor do mundo.

5.2.1. Estados Unidos

A Fundação Nacional de Ciências norte-americana (*National Science Foundation – NSF*) foi um dos primeiros financiadores públicos a avaliar o impacto social da pesquisa, e o potencial avanço na área do conhecimento, isso na década de 90 (HOLBROOK; FRODEMAN, 2010, p. 218 apud BORNMANN, 2013b). O Conselho Nacional de Ciência (*National Science Board – NSB*) adicionou em 2011, critérios de análise social no investimento realizado, incluindo critérios nacionais como benefícios potenciais da pesquisa à nação norte americana, bem como potenciais inovações.

Atualmente, a NSF exige, em cada projeto, uma seção detalhando o impacto social da ciência, mas não mais define o que seria esse impacto. Os revisores são orientados a examinar esse impacto e dar a sua devida importância. O seguinte trecho mostra como essa seção deve ser elaborada:

A descrição do projeto deve conter uma seção separada chamada "impactos mais amplos". Esta seção deve fornecer uma discussão sobre os impactos mais amplos das atividades propostas. Impactos mais amplos podem ser conseguidos através da própria pesquisa, por meio das atividades que estão diretamente relacionadas com projetos de investigação específicos, ou por meio de atividades que são suportadas pelo projeto, mas são complementares. A NSF valoriza o avanço do conhecimento científico e atividades que contribuem para a obtenção de resultados socialmente relevantes. Esses resultados incluem, mas não estão limitados a: plena participação das mulheres, pessoas com deficiência e minorias sub-representadas na ciência, tecnologia, engenharia e matemática (science, technology, engineering, and mathematics - STEM); melhoria da educação e dos educadores em STEM; aumento da literacia científica e envolvimento do público com a ciência e tecnologia; melhoria do bem-estar dos indivíduos na sociedade; desenvolvimento de uma mão de obra diversificada em STEM, competitiva em nível mundial; o aumento das parcerias entre o meio acadêmico, a indústria, e outros; melhoria da segurança nacional; aumento da competitividade econômica dos EUA; e infraestrutura para pesquisa e educação reforçada.

(“Proposal and Award Policies and Procedures Guide, January 2016”, 2016).

Dessa forma, o modelo norte-americano evoluiu de tal forma a permitir que seja o próprio pesquisador que enfatize eventual impacto social, econômico, político ou ambiental de sua pesquisa, impacto esse a ser eventualmente avaliado por um comitê especialista na área.

5.2.2. Austrália

O Grupo Consultivo de Desenvolvimento (Development Advisory Group – DAG) foi criado em 2006 com a tarefa de implementar a Análise da Qualidade da Pesquisa (Research Quality Framework – RQF), que foi proposta como ferramenta para análise da qualidade científica, bem como eventuais impactos sociais, econômicos, políticos e ambientais, buscando ligar as pesquisas à indústria e ao comércio, buscando um maior benefício social. Assim, o DAG estabeleceu o *Technical Working Group on Research Impact* (Grupo Técnico para o Impacto da Ciência), com pesquisadores das várias áreas do conhecimento e *expertise* no assunto, além de representantes da indústria, do comércio e do governo, com o objetivo de desenvolver uma metodologia e indicadores para avaliar o impacto, definido de modo amplo (DONOVAN, 2008).

Assim, o RQF seria formado de 13 painéis, um para cada área do conhecimento, que avaliaria a qualidade por meio de quatro itens representativos do sucesso do pesquisador, com indicadores subsidiando a questão. Dessa forma, a qualidade do pesquisador seria descrita pelos seus quatro melhores trabalhos, ou suas quatro maiores conquistas. Foi então criada uma escala do impacto, com participação de uma autoavaliação realizada pelo pesquisador, que indicaria não só os impactos sociais, econômicos, políticos e ambientais, mas também os beneficiados e as partes interessadas (DONOVAN, 2008).

No entanto, o grupo chegou à conclusão de que os indicadores existentes não são adequados para representar o impacto, embora possam ser usados para corroborar o impacto, apesar de que novos indicadores poderiam ser eventualmente desenvolvidos. Assim, a conclusão foi de que há incompatibilidade entre análise de impacto social e econômico, cujos indicadores seriam investimentos na indústria, patentes e comercialização, com os valores intangíveis resultantes da pesquisa científica, que somente podem ser analisadas por um corpo consultivo especializado (por pares). Assim, a análise por pares foi a adotada pelo DAG, por meio de estudos de caso e, embora todo o desenho tenha sido realizado para que fosse medido o impacto da ciência em todas as suas dimensões, o resultado final tendeu a uma análise da qualidade da pesquisa em si, numa abordagem qualitativa e contextual, por meio de revisão por

pares baseados em indicadores de qualidade da ciência, culminando na descontinuidade do RQF em 2007 (DONOVAN, 2008).

5.2.3. Países Baixos

Os Países Baixos também buscam medir a qualidade da pesquisa por meio dos pares especialistas na área do projeto científico em questão, com base no *Standard Evaluation Protocol* (SEP), documento que busca guiar a forma como a avaliação deve ser realizada. O SEP se constitui, na verdade, de um documento político mais abrangente, que busca: orientar os pesquisadores sobre como avaliar a qualidade da pesquisa, sua relevância social e as unidades estratégicas; orientar os dirigentes a medir o impacto de suas políticas públicas; orientar o governo a usar essas análises e a promover um sistema de pesquisa de excelência; e orientar a sociedade e o setor privado, por serem partes interessadas nas pesquisas e nos avanços científicos (VSNU; OTHERS, 2009).

Dessa forma, uma ampla avaliação é realizada a cada 6 anos e é de responsabilidade das próprias instituições, que definem as linhas gerais dessa análise, sua frequência e sua publicidade. A análise é realizada por um comitê imparcial, preferencialmente internacional, com o uso de autoavaliações, entrevistas e outros documentos, de modo a aferir a qualidade da pesquisa (com o uso de indicadores como o número de publicações, e infraestrutura), a relevância social (contribuições a grupos econômicos, sociais ou culturais) e a viabilidade (estratégia e capacidade de atingir metas previstas nos próximos anos). São exemplos de indicadores usados: artigos, livros, produção técnica, teses, citações, laboratórios, premiações, participações em congressos, relatórios (por exemplo, para formação de política pública), artigos em jornais e revistas, infraestrutura, softwares, palestras, feiras, patentes, compartilhamento de conhecimento e infraestrutura, colaborações, inclusive com grupos sociais, etc. Para o autor, a Holanda possui um dos mais evoluídos exemplos de avaliação do impacto que é conduzida lado a lado com a análise da qualidade científica (DONOVAN, 2008).

Der Meulen et al. (2000) elogiam a forma como os Países Baixos analisam o impacto da pesquisa e consideram que o modelo é adequado para ser aplicado em outros países, sendo que cerca de 80% das avaliações realizadas incluem análise de impacto com a expectativa de que o modelo adotado traga real benefício ao país (DER MEULEN; RIP, 2000; DONOVAN, 2008).

5.2.4. Comissão Europeia (E3M)

Oito países europeus financiaram um projeto de três anos no âmbito do Programa de Aprendizagem ao Longo da Vida (*Lifelong Learning Programme*) da União Europeia, denominado de E3M, com o objetivo de promover uma forma de avaliação da terceira missão das universidades, ou seja, seu impacto social. O projeto buscou indicadores e ferramentas para promover essas análises, bem como o desenvolvimento de uma rede internacional com conferências e estudos de caso. Os resultados encontrados foram publicados sob o nome de *Green Paper* (SOEIRO et al., 2012). Pelo estudo, indicadores e métricas são úteis e podem dar informações relevantes, mas devem ser usados com cuidado, pois os indicadores não fornecem resultados absolutos ou verdadeiros e sim uma representação da realidade. Além disso, o uso de indicadores provoca distorções, na medida em que pesquisadores buscarão a manipulação do sistema e o aumento do indicador sem aumentar a qualidade da ciência produzida. Ainda assim, indicadores podem ser utilizados para subsidiar a tomada de decisões, bem como para encontrar trabalhos relevantes, antes ocultos, ou atrair investimentos (SOEIRO; MONTESINOS; OTHERS, 2011).

Ainda assim, há dificuldades em se medir ou identificar a terceira missão, sendo que dezenas de indicadores são necessários para medir o impacto de forma rica e fiel. Mas, se a intenção for apenas medir ou classificar o grau de enquadramento das instituições na terceira missão, apenas três ou quatro indicadores são necessários, conforme cada caso, muito embora esses variem conforme o caso e a escolha de quais seriam esses indicadores seja uma tarefa complicada em si. O método usado pelo E3M é a aplicação de três rodadas de questionários com cerca de 95 indicadores representando a educação, a transferência de tecnologia e o engajamento social, na primeira rodada, 63 indicadores na segunda rodada e 24 indicadores na terceira (SOEIRO; MONTESINOS; OTHERS, 2011).

5.2.5. Alemanha

O fomento à pesquisa na Alemanha é realizado de modo descentralizado, com agências federais e estaduais, bem como com instituições tanto públicas quanto privadas (MÜNCH; SCHÄFER, 2014). São 38 institutos financiados pelo governo alemão, cada um relacionado a uma atividade e oferecendo base científica para a execução de projetos. Para avaliar a pesquisa foi criada a Iniciativa Alemã de Excelência (*German Excellence Initiative*) em 2006, que orienta quais instituições financiar. São três distintas áreas de financiamento: (1) escolas de pesquisa;

(2) universidades de excelência; e (3) Colaborações de Excelência (*Clusters of Excellence*) (MÖLLER; SCHMIDT; HORNBOSTEL, 2016).

A Sociedade Alemã de Amparo à Pesquisa (DFG) é a maior organização de fomento de pesquisa alemã e seu processo de seleção inclui revisão por comitês científicos, especialmente escolhidos e sem retribuição financeira, que analisa cada proposta por seu mérito científico (FORSCHUNGSGEMEINSCHAFT, 2012). São 48 comitês ao todo, livres para estabelecer suas próprias regras de avaliação, eleitos para um mandato de 4 anos, com no mínimo dois representantes por área (por comitê) (FORSCHUNGSGEMEINSCHAFT, 2012).

5.2.6. Brasil – MCTI

No Brasil, conforme o [Decreto nº 11.493, de 17 de abril de 2023](#), compete ao Ministério da Ciência e Tecnologia (MCTI) os assuntos relacionados às políticas e os relacionados à coordenação, supervisão, monitoramento e avaliação das atividades nacionais de ciência, tecnologia e inovação. Dentro do MCTI, compete à Secretaria-Executiva “articular, propor e implementar mecanismos institucionais de prospecção e monitoramento da evolução do progresso científico e tecnológico no País e no exterior, em especial em áreas de interesse estratégico para o desenvolvimento nacional”.

Como parte de suas funções, o MCTI acompanha diversos indicadores de CT&I. Como exemplo, o MCTI lançou recentemente o documento [Indicadores Nacionais de Ciência, Tecnologia e Inovação 2022](#). No entanto, observa-se que os indicadores monitorados são todos de insumo, como dispêndio realizado ou renúncia fiscal autorizada. Uma análise rápida do documento nos permite inferir que:

- Houve aumento do investimento em CT&I até 2015 e redução do investimento a partir de 2016.
- Investimos metade do que deveríamos, se comparado a países desenvolvidos e em relação ao PIB.
 - Brasil se equivale a Rússia e Índia em % do PIB investido. Coréia investe quatro vezes mais, EUA três.
 - Brasil mantém (reduziu a partir de 2017) o investimento em relação ao PIB. Países de primeiro mundo tenderam ao aumento.
- O Ministério da Educação investe duas vezes mais em ciência e tecnologia que o MCTI, o que pode ser indicador de desvio de finalidade.

- 65,6% dos investimentos são no Sudeste, sendo mais de 53% em SP. Há falta de equidade regional. Mas, por outro lado, SP é o único estado que investe porcentagem de seu PIB, compatível a de países de primeiro mundo, em CT&I.
- Há pouca presença de pesquisadores na indústria ou no comércio, pois o ensino superior concentra quase cinco vezes mais pesquisadores que todos os outros setores.
- Há mais pesquisadores Mestres que Doutores.
- Há redução na formação humana nos últimos anos (graduados, mestres e doutores).
- Há alta concentração de pesquisadores em Ciências Humanas (área não prioritária) e na Saúde. Ou seja, há falta de equidade na distribuição entre as diversas áreas do conhecimento.
- Muitos indicadores só foram medidos até o ano de 2016, o que impede uma análise detalhada ou atual.
- Há redução de bolsas de pós-graduação no exterior. No entanto, isso decorre, provavelmente, devido ao fim do programa Ciência sem Fronteiras.
- Houve congelamento do número de bolsas a partir de 2017.
- Houve permanência em menos de 3% do percentual de artigos no mundo.
- Há forte redução no número de artigos indexados pela Scopus (mantendo o percentual de citação). No entanto, isso pode ser devido ao incremento de revistas indexadas no sistema e o indicador pode ser elusivo.
- Houve redução no número de pedidos de patentes, mas crescimento exponencial no número de concessões. Assim, acredita-se que haja a presença de algum fator oculto que precisa ser estudado.
- Houve aumento de pedidos e concessões de patentes no exterior. Esse é um bom indicador de potencial inovador de alcance internacional.
- No Brasil, o maior investimento é pelo setor público. Em países do primeiro mundo, o maior investidor são empresas. Vale ressaltar que estudos indicam que o investimento em CT&I realizado por empresas são mais eficientes para alavancar o crescimento do país que os realizados pelo setor público.
- Uma curiosidade: o aumento da China em artigos publicados é excepcional.

O MCTI também lançou o [Plano Estratégico do Ministério da Ciência, Tecnologia e Inovações - PEI MCTI 2020-2023](#). Nele, constitui missão do MCTI “produzir conhecimento,

produzir riqueza para o Brasil e contribuir para a qualidade de vida dos brasileiros” (MCTI, 2022). O documento apresenta o Mapa Estratégico 2020-2030, estabelece 15 objetivos a alcançar, sendo o primeiro, para exemplificar, “Ampliar e fortalecer a capacidade científica, tecnológica e a infraestrutura de pesquisa e desenvolvimento do país” e estabelece indicadores para cada um desses objetivos. Observa-se que os indicadores podem ser separados nos seguintes grupos:

- Produtividade (publicações, patentes);
- Qualidade (citações);
- Insumo (investimento);
- Quantitativos (nº de beneficiários, supressão vegetativa, nº de acordo, etc.);
- Infraestrutura (nº de ICTs, escolas, etc.);
- Regionalidade (municípios ou estados atendidos);

Assim, os indicadores usados são os mesmos comumente utilizados pelo restante do mundo. No entanto, dentre os indicadores propostos, um se destaca para este estudo: “Indicador 03: Taxa de políticas de CT&I avaliadas. Este indicador reflete o esforço necessário para avaliar e incrementar a eficiência, a eficácia e a efetividade das políticas públicas de ciência, tecnologia e inovação”. O indicador se insere no objetivo 8: “Promover alternativas ao orçamento público para o fomento de CT&I”. Observa-se, no entanto, que uma pesquisa no Google por "Taxa de políticas de CT&I avaliadas" não encontrou qualquer resultado. Tão pouco foi encontrada qualquer pesquisa, dentro do MCTI, de impacto de política pública de CT&I nos moldes desta pesquisa. Finalmente, foi aberta demanda pelo portal de informações do governo federal solicitando mais informações, ainda sem resposta.

5.2.7. Brasil – CGEE

O Centro de Gestão e Estudos Estratégicos (CGEE) é uma associação civil sem fins lucrativos, que tem, entre seus [objetivos estatutários](#), “promover e realizar atividades de avaliação de estratégias e de impactos econômicos e sociais das políticas, dos programas e projetos científicos, tecnológicos, de inovação e de formação de recursos humanos”⁹.

Filiado ao CGEE há o [Observatório de Ciência, Tecnologia e Inovação](#) (OCTI), que “tem como objetivo principal o monitoramento da produção científica, tecnológica e da inovação, acompanhando tendências no Brasil e no mundo. O OCTI elabora, ainda, indicadores

⁹ <https://www.cgee.org.br/missao-e-objetivos>

que permitam orientar e monitorar políticas públicas de CT&I no país e avaliá-los no contexto internacional”¹⁰.

A CGEE apresenta estudos sobre o progresso da CT&I no Brasil, com o uso de diversos indicadores mundialmente reconhecidos. Como exemplo, o Boletim Anual OCTI 2021, Ano 2 – maio de 2022, dá fiel retrato da ciência brasileira por meio de análise dos artigos publicados, o ano, a área, taxas de crescimento, região, comparação com outros países, análise de clusters temáticos, nuvens de palavras e gráfico de radar com indicadores de formação e inovação.

No entanto, apesar de um dos objetivos ser justamente encontrar o impacto econômico da política realizada, não foram encontrados estudos de impacto que abranjam as dimensões sociais, culturais, econômicas e ambientais. Os estudos, quando presentes, seguem a tendência mundial de medir a produtividade (artigos publicados, por exemplo), qualidade (representada, principalmente, pelo número de citações recebidas) e colaboração (análise dos autores, como se relacionam e suas instituições). Assim, muito menos foi identificada qualquer publicação por parte do CGEE com o uso de um modelo econométrico que busque inferir a causalidade de uma política pública usando o Currículo Lattes como ferramenta bibliométrica.

5.2.8. Brasil – CNPq

O CNPq adota metodologia internacionalmente aceita para a escolha dos projetos de pesquisa a serem financiados, sendo que a forma como realiza o julgamento de quais projetos de pesquisa serão ou não financiados se aproxima do modelo alemão. Os recursos a serem aplicados em CT&I pelo CNPq são repassados pelo Governo Federal de forma direta (transferência direta de recursos) e indireta (por meio de transferência proveniente de outros órgãos e fundos). Independentemente da origem dos recursos, é estabelecido o valor a ser financiado e o objetivo do financiamento. Com base no objetivo, um edital público é lançado.

Dessa forma, o CNPq trabalha com dezenas (ou até centenas) de editais públicos, muitos temáticos. No edital são estabelecidas as regras da concorrência, o calendário, como será realizado o julgamento, os benefícios que os contemplados terão, enfim, toda a regra do negócio, que se torna um documento público. Com base no edital, cada pesquisador deve submeter seu projeto de pesquisa para análise, sendo que o correto preenchimento do Currículo Lattes é obrigatório.

¹⁰ <https://octi.cgge.org.br/sobre>

Cada projeto passa por uma pré-seleção técnica, onde são analisados os quesitos técnicos e garantida a correta concorrência pública, como também passa por análises pelos pares na forma de pareceres individuais (Ad-Hocs). Para cada área de pesquisa há um Comitê de Assessoramento (CA).

O Comitê de Assessoramento (CA) é eleito e formado por especialistas da área, que já passaram por uma ampla e concorrida seleção pública e que, portanto, tiveram sua excelência e qualidade na pesquisa reconhecidos. Essa seleção pública é a concorrência a uma Bolsa Produtividade, que, além de pagar uma pequena complementação salarial mensal, se tornou um prêmio de reconhecimento pela relevância científica do pesquisador, altamente valorizado e concorrido.

Dentre os mais reconhecidos desses pesquisadores de cada área há uma eleição de modo a formar lista tríplice, que é enviada ao Conselho Deliberativo do órgão para a escolha o membro do CA, que possuirá mandato de dois anos, renováveis uma vez. São 49 CAs ao todo, um para cada área de conhecimento, sendo que cada um possui uma certa quantidade de membros que representam toda a comunidade científica e estão entre os mais prestigiados e reconhecidos pesquisadores da área.

Cabe ao CA realizar a classificação das propostas quanto ao mérito científico, de forma que cada projeto é então concedido e financiado se houver os recursos para isso e se o mérito da proposta tiver sido reconhecido pelo CA. Dessa forma, em termos da escolha dos projetos que serão aprovados, são os CAs os órgãos deliberativos de maior importância.

Os projetos são, normalmente, de livre apresentação e não há modelo previamente estabelecido, de modo que não há exigência de que o projeto inclua os impactos sociais da pesquisa. Apesar disso, a chamada pública pode estabelecer um modelo em que os impactos são exigidos e, mesmo na ausência desse modelo, o impacto social é entendido como relevante e há o incentivo que seja incluso.

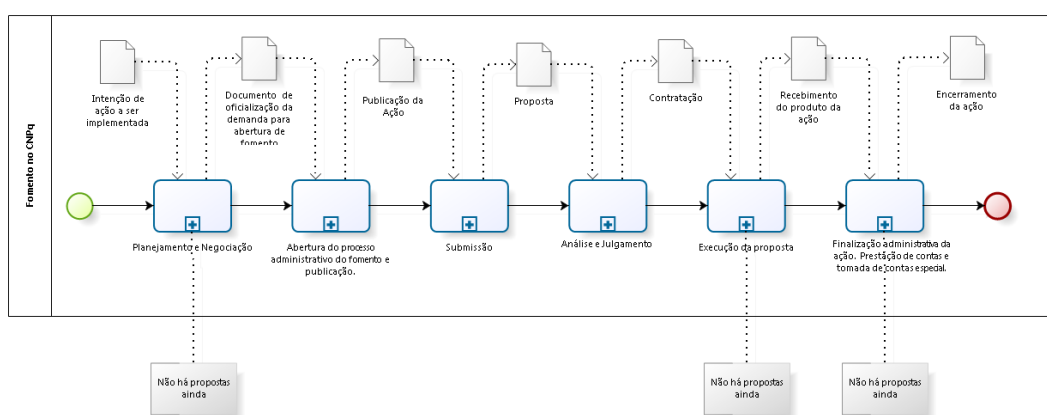
Cada CA decide sobre quais indicadores utilizar e qual a metodologia adotada, com base nos critérios de julgamento estabelecidos na Chamada. Dessa forma, cada área de pesquisa possui um julgamento e uma classificação que depende de análise pelos pares, subsidiada pelo Currículo Lattes, o projeto de pesquisa apresentado e uma planilha com indicadores. A Chamada ainda pode prever Comitês Especiais, formados por pesquisadores especialistas convidados, nacionais ou estrangeiros, para análise, classificação as propostas, visitas *in loco*, congressos e apresentação de resultados à sociedade.

Esse modelo de financiamento pode ser substituído por parcerias com outras instituições. Por exemplo, no caso de bolsas de doutorado, cada Programa de Pós-Graduação

(PPG) possui uma certa quantidade de bolsas que podem ser por ele livremente distribuídas, tendo regras pré-estabelecidas que impedem o mau uso dos recursos. Dessa forma, cabe a cada PPG determinar como será dada essa concorrência interna às bolsas disponíveis. Esse modelo descentralizado permite que a própria instituição se utilize das bolsas da forma o mais eficiente o possível, sendo que cada concessão é fiscalizada pelo CNPq.

Cada concessão, seja ela no formato de bolsas, seja ela no formato de financiamento a pesquisas, é acompanhada anualmente, por meio de relatórios, bem como ao final, por meio de um Relatório Técnico Final, que é analisado pelos técnicos do CNPq. Dessa forma, qualquer desvio de finalidade que for encontrada é punido e os recursos devolvidos ao governo federal.

Figura 14 – Fluxo do Fomento do CNPq

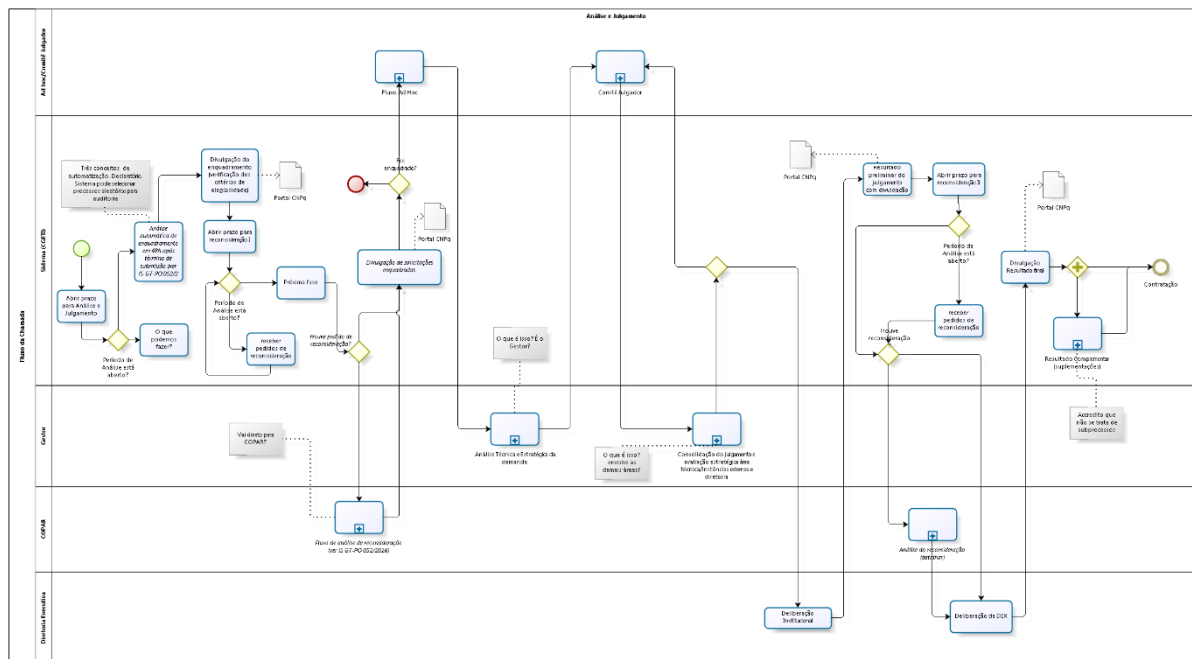


Fonte: Autoria Própria.

Apesar desse rigoroso processo de seleção, não há atualmente, no CNPq, qualquer fluxo de análise de impacto de suas políticas públicas instituído. Ou seja, muito embora cada projeto seja seguido de perto e o mérito de cada proposta rigorosamente julgado antes de que a concessão seja realizada, não há estudos sistematizados que indiquem os impactos dos financiamentos realizados.

Assim, as políticas públicas são normalmente concedidas com base no histórico de “sucessos” e “fracassos”, sendo o sucesso ou o fracasso medido de modo informal. Apesar disso, o CNPq é tido como uma das principais agências de fomento científico dos países em desenvolvimento, considerado como um modelo a ser seguido ou adotado.

Figura 15 – Fluxo de Análise e Julgamento das Propostas



Fonte: Autoria Própria.

5.2.9. Conclusão

Observa-se que o E3M da Comissão Europeia chegou na conclusão de que a análise da terceira missão é possível com o uso de dezenas (ou mesmo centenas) de indicadores, algo não facilmente atingível sem uma quantidade considerável de trabalho ou de recursos. No entanto, a dificuldade observada pode ser derivada do fato de que não foi encontrado, no levantamento bibliográfico realizado, instituição com ferramenta similar ao Currículo Lattes adotado pelo CNPq, que possui mais de uma centena de indicadores, o que pode possibilitar a análise da terceira missão.

Ainda assim, o caso da Austrália indica que mesmo que diversos indicadores sejam usados, esses nunca poderão ser substituídos por uma análise pelos pares, por especialistas naquela área do conhecimento. Dessa forma, uma análise puramente quantitativa pode, no máximo, servir como base para a avaliação, o que reforça que o impacto alcançado por este estudo não pode ser usado como ferramenta direta de escolha de política pública.

Ainda assim, a Comissão Européia demonstra como indicadores são úteis e que podem ser utilizados para subsidiar a tomada de decisões, bem como para encontrar trabalhos relevantes, antes ocultos, ou atrair investimentos, mesmo tendo a desvantagem de poder causar o efeito perverso de ter pesquisadores buscando o acréscimo do indicador, e não a qualidade da

pesquisa, e mesmo o indicador apenas representar (e não espelhar) uma pequena parte da realidade.

Dessa forma, acredita-se que uma das melhores formas de análise do potencial impacto da pesquisa seja o norte-americano, em que o próprio pesquisador indica esse impacto, que é avaliado pelos pares. O modelo alemão é especialmente útil para indicar como organizar essa análise pelos pares e se aproxima da realidade do CNPq, sendo que o modelo australiano demonstrou que o impacto social da ciência, potencial ou realizado, não pode ser medido por indicadores previamente estabelecidos. O potencial impacto não pode ser medido por indicadores, mas o impacto realizado pode ser corroborado por esses.

Dessa forma, o cálculo do impacto da pesquisa pode permitir a comparação entre políticas públicas de fomento, sendo que não foi encontrada, na literatura mundial levantada, qualquer avaliação sistemática, longitudinal e horizontal, recorrente e imparcial do potencial impacto do investimento realizado pelas agências públicas de financiamento à pesquisa.

O cálculo pode incluir a terceira missão das instituições e pode ser útil para permitir análises e subsidiar decisões, mas não pode ser diretamente adotado para escolha de políticas públicas ou escolha de projetos a serem financiados sem antes passar pelo crivo de analistas ou especialistas no assunto. Também, observa-se que não se pode usar indicadores para encontrar o impacto em modelo preditivo (buscar qual será o futuro impacto de um projeto apresentado), mas apenas corroborar o impacto sugerido ou já causado.

5.3. Escolha das Políticas Públicas a serem analisadas

Para a escolha de duas políticas públicas a serem analisadas neste estudo, verificou-se quais os tipos de financiamentos realizados CNPq. Desde a sua informatização, o órgão registrou 975 financiamentos do tipo “Edital” e 549 do tipo “Bolsas e Auxílios”. Os registros podem ser separados em grupos, como financiamento a Institutos de Tecnologia, concessão de bolsas Produtividade, o Edital Universal e a concessão de bolsas de Mestrado e Doutorado (GM e GD respectivamente).

Desses grupos, os Editais Universais representaram 14,36% da demanda total, os Institutos Nacionais de Ciência e Tecnologia representam 11,65% e Bolsas Produtividade representam 10,04% dessa demanda. Mas, se considerarmos o valor efetivamente pago até o ano de 2022 (ao invés da demanda bruta), de acordo com o Portal de Dados Abertos do CNPq¹¹,

¹¹ Constatou-se que o Portal de Dados Abertos do CNPq não possui a completude dos dados, principalmente nos anos iniciais, em que não há dados de pagamentos a Projetos de Pesquisa, somente a bolsas.

Bolsas Produtividade (PQ) representam 17% dos pagamentos realizados, 15% de Doutorado (GD) e 9% de Mestrado (GM). Dessa forma, observa-se que bolsas de doutorado receberam quase a mesma quantidade de recursos que Bolsas Produtividade.

A bolsa Produtividade em Pesquisa é “Destinada aos pesquisadores que se destaquem entre seus pares, valorizando sua produção científica (...)” (CNPQ, 2015). Assim, essa bolsa é, em certa medida, uma “bolsa prêmio”, que não é concedida de modo a financiar um projeto de pesquisa, mas um reconhecimento do mérito científico do pesquisador. Dessa forma, o impacto da bolsa PQ é difuso e não se concentra apenas no currículo do pesquisador. Além disso, foram encontrados diversos estudos relacionados à bolsa PQ do CNPq, de modo que um estudo nesse sentido seria de pouca ajuda para o avanço na fronteira do conhecimento.

Já o Programa “Institutos Nacionais de Ciência e Tecnologia” tem por objetivo apoiar atividades de pesquisa de alto impacto científico em áreas estratégicas (CNPQ, 2014). Ocorre que são 101 os INCTs, número que dificulta a aplicação direta de um modelo econométrico, dificultando o uso desse financiamento neste estudo. Além disso, a concessão a INCTs é destinada a um coordenador e o impacto causado é muito maior do que pode ser representado apenas pelo seu currículo e, portanto, também difuso. Assim, a princípio, recomendar-se-ia estudos de casos para a análise do impacto ao financiamento a INCTs. Ou seja, para o uso do modelo econométrico a ser proposto em um INCT seria necessário um estudo mais rigoroso, que verificasse o impacto não só no Coordenador do Projeto, e não só os colaboradores, mas em toda a instituição e talvez até sociedade.

O Edital Universal tem como objetivo “selecionar propostas para apoio financeiro a projetos que visem contribuir significativamente para o desenvolvimento científico, tecnológico e de inovação do País, em qualquer área do conhecimento” (CNPQ, 2010). O Universal possui objetivo de financiamento a pesquisas, é aberto a todas as áreas e financia tanto pequenas quanto grandes pesquisas em todo o Brasil. Além disso, a forte concorrência e o número de financiamentos realizados permitem a aplicação direta de um modelo econométrico, sendo que o impacto se concentra na figura do pesquisador financiado e não em toda uma instituição.

Da mesma forma, o alto quantitativo de bolsas de doutorado, sem que tenha sido encontrada na literatura análise do impacto dessas por meio de um modelo econométrico, torna essas como alvo ideal deste estudo. O número de bolsas concedidas é alto, o que permite o uso do modelo econométrico, e o impacto também concentrado na figura do pesquisador financiado (e não em toda a instituição).

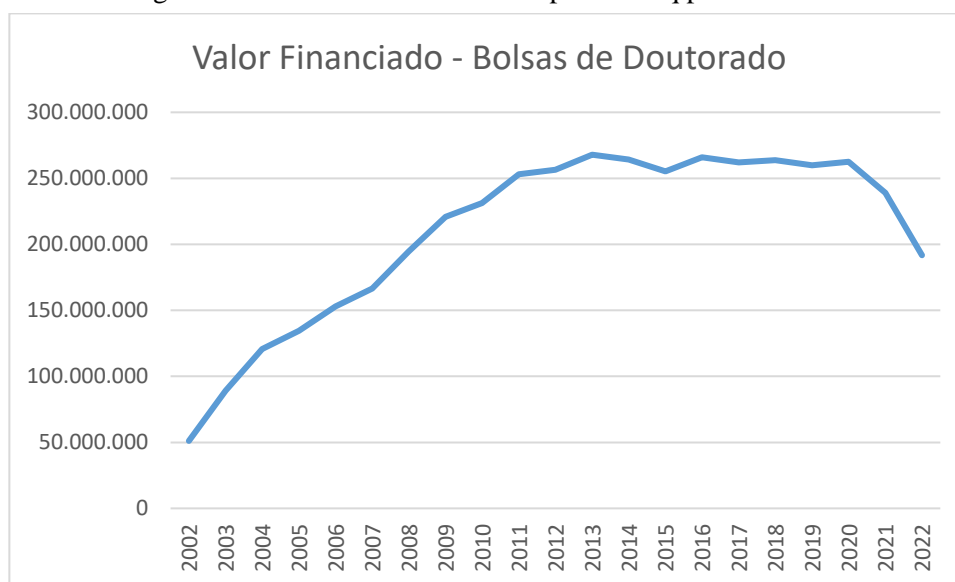
Assim, o alto volume de recursos financiados, o alto número de pesquisadores contemplados, o fato de que se espera que o impacto se concentre, em certa medida, nos pesquisadores contemplados, tornam a Concessão de Bolsas de Doutorado (bolsas GD), bem como o Edital Universal, ideais para este estudo.

5.3.1. Bolsas de Doutorado

O CNPq concede bolsas de doutorado por meio de editais públicos em que participam representantes de PPGs, que recebem cotas de bolsas de Doutorado (bolsas GD) e podem distribuir essas cotas internamente entre seus alunos. Para se candidatar a uma bolsa GD é necessário atender a certos critérios, como ter dedicação integral às atividades acadêmicas, estar devidamente matriculado no PPG, possuir um projeto de pesquisa e demonstrar um bom desempenho acadêmico. Muitas instituições promovem seleções internas com regras próprias, permitindo uma concorrência ampla e justa. Além disso, o CNPq instituiu regras para a distribuição dessas bolsas pelas instituições, o que impede o mal uso dos recursos públicos.

A bolsa GD tem duração de até 48 meses e possui o valor atual de R\$ 3.100,00 mensais, de modo que o valor total financiado pode chegar a R\$ 148.800,00. O valor, longe de ser um alto salário, é considerado apenas o suficiente para a manutenção do estudante em tempo integral para seu estudo.

Figura 16 – Valor Total Concedido pelo CNPq para Bolsas GD



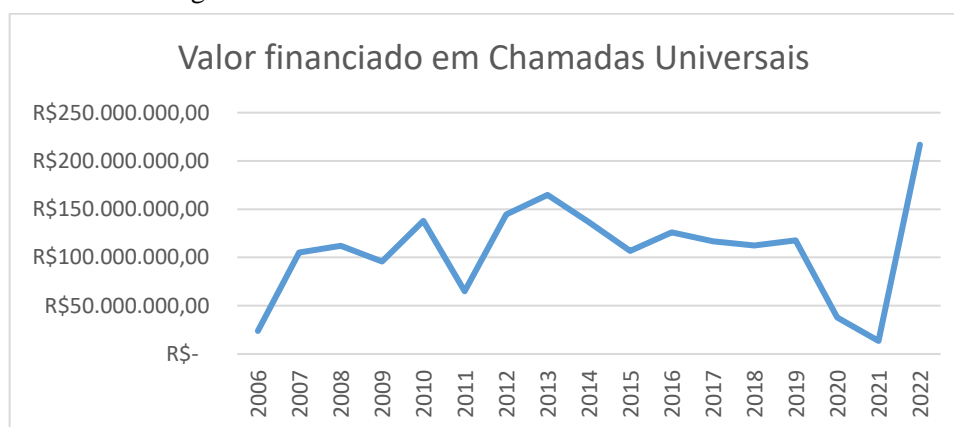
Fonte: Autoria Própria.

5.3.2. A Chamada Universal

O primeiro Universal foi lançado em 1999 e, desde então, é uma chamada pública que destina mais de cem milhões de reais, anualmente, a milhares de projetos de pesquisa em todas as áreas do conhecimento. Os recursos são divididos quanto às prioridades governamentais e de forma a distribuir conforma a demanda bruta encontrada, de modo que, quanto maior a demanda em determinada área, mais recursos essa área receberá¹².

O Universal é um dos “carros chefes” do CNPq e pode financiar qualquer pesquisa científica, tecnológica ou de inovação que tiver seu mérito científico reconhecido no processo de julgamento, sem restrições temáticas. O Universal disponibiliza recursos financeiros que podem ser utilizados para financiar despesas como contratação de pessoal, aquisição de equipamentos e materiais, viagens de estudo e outros gastos relacionados à execução do projeto, sem necessitar, para isso, obedecer aos trâmites legais de licitações públicas. Assim, o Universal “oxigena” a ciência brasileira, permitindo que o pesquisador consiga aquilo que é necessário para sua pesquisa, de forma simples rápida e direta.

Figura 17 – Valor financiado em Chamadas Universais



Fonte: Autoria própria.

A concorrência é realizada por meio do envio de um projeto de pesquisa no *site* do CNPq e aberto a qualquer pesquisador. O CNPq realiza chamadas anuais, publicando o edital em sua página. Nele, o pesquisador pode ter instruções claras de como concorrer e as regras da concorrência, que são simples e diretas. Não há modelo de projeto de pesquisa, mas a concorrência é forte, de modo que apenas os melhores projetos são selecionados. Normalmente são mais de 10.000 propostas concorrentes, dessas apenas cerca de 20% serão escolhidas para financiamento.

¹² <https://youtu.be/ffP8xcENrQY>

O processo de avaliação é uma sequência que envolve uma análise pelo corpo técnico do CNPq, onde verifica-se se a proposta se enquadra dentro das regras do concurso. Depois, a proposta é analisada por pares (Ad-Hocs), que são pesquisadores PQ¹³. Finalmente, um CA¹⁴ tem acesso ao conjunto de propostas de sua área e realizam, com base nos pareceres emitidos, nos projetos de pesquisa e nos Currículo Lattes dos pesquisadores, a classificação das propostas.

As regras de classificação são previamente estabelecidas, diferentes para cada área do conhecimento, publicamente disponíveis e claras. Dessa forma, as regras permitem um julgamento transparente e justo. Por outro lado, o julgamento é normalmente subjetivo, uma vez que o mérito da proposta passa pelo julgamento pessoal dos especialistas em questão. De qualquer forma, em busca de garantir a transparência do processo, é emitido um parecer em cada etapa do julgamento, disponível ao pesquisador, que indica a classificação alcançada, bem como os pontos meritórios ou não.

5.3.3. Conclusão

A importância de acompanhar e analisar a CT&I de um país é, portanto, mundialmente reconhecida e há muitas iniciativas nesse sentido. No entanto, observa-se pouca presença em análise de impacto de um investimento público em CT&I. Essa, quando é feita, na maioria das vezes usa apenas de indicadores facilmente obtidos, como o número de artigos publicados, o número de citações, os autores envolvidos (e as respectivas redes de colaboração).

Analisando a literatura mundial, não foi encontrado estudo capaz de inferir a correlação entre o financiamento a um projeto de pesquisa e a formação de um doutor, uma publicação em um jornal ou revista de grande circulação, uma obra de arte ou mesmo uma premiação obtida. Um dos possíveis motivos é a ausência de um instrumento como o Currículo Lattes, que apresente todos esses indicadores e que, portanto, permite a realização dessa inferência.

5.4. Método de levantamento bibliográfico

A pesquisa cientométrica é interdisciplinar na temática ou multidisciplinar na origem e permeia diversas áreas da ciência (TEIXEIRA et al., 2018). Por isso, realizar o levantamento bibliográfico é um desafio. Uma busca por “Impacto Ciência” no Portal de Periódicos da

¹³ PQ – Produtividade em Pesquisa. É um edital altamente concorrido que premia os melhores pesquisadores do país com bolsas e/ou adicionais para uso em pesquisas. Ser bolsista PQ é tido como certificação de excelência.

¹⁴ CA – Comitê Assessor. Comitê realizado pelos pesquisadores PQ de maior nível, publicamente indicados e escolhidos pelo órgão máximo do CNPq, o Conselho Deliberativo.

CAPES revelou mais de 40.000 materiais a respeito, incluindo 21.260 artigos revisados por pares. Ou seja, o assunto é por demais abrangente e precisa ser refinado.

Para o levantamento bibliográfico, precisamos saber que o estudo cientométrico e bibliométrico busca a análise de uma política pública de investimento à ciência com o uso de *Data Mining* no Currículo Lattes com o uso da ciência de dados para coleta de indicadores a serem aplicados em uma análise econométrica de forma a calcular o impacto dessa política pública através de um quase-experimento com o método das diferenças em diferenças:

- **Cientometria:** ramo da sociologia das ciências e da ciência da informação que procura estudar aspectos quantitativos da ciência e da produção científica, quer como uma disciplina, quer como uma atividade econômica¹⁵;
- **Bibliometria:** parte da cientometria, é a técnica de análise de pesquisa que estuda publicações em livros, relatórios e artigos para quantificar, analisar e avaliar a produção acadêmica científica de temas¹⁶;
- **Data mining:** processo de explorar grandes quantidades de dados à procura de padrões consistentes, com o uso de ferramentas capazes de explorar um conjunto de dados, extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento¹⁷;
- **Ciência de dados:** Ramo da *Data Mining*, estuda os dados para extrair *insights* significativos para os negócios. Ela é uma abordagem multidisciplinar que combina princípios e práticas das áreas de matemática, estatística, inteligência artificial e engenharia da computação para analisar grandes quantidades de dados. Essa análise ajuda os cientistas de dados a fazer e responder perguntas como o que aconteceu, por que aconteceu, o que acontecerá e o que pode ser feito com os resultados¹⁸.
- **Currículo Lattes:** um sistema de currículos virtual criado e mantido pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo qual

¹⁵ <https://pt.wikipedia.org/wiki/Cienciometria>

¹⁶ http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1562-47302017000400001#:~:text=A%20bibliometria%20%C3%A9%20uma%20t%C3%A9cnica,ao%20objetivo%20geral%20deste%20estudo.

¹⁷ https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjZsdKik_D-AhXmBLkGHdb_DR0QFnoECA0QAw&url=https%3A%2F%2Fcetax.com.br%2Fdata-mining%2F%23%3A~%3Atext%3DConsideramos%2520Data%2520Mining%2520ou%2520Minera%25C3%25A7%25C3%25A3o%2Cassim%2520novos%2520subconjuntos%2520de%2520dados.&usg=AOvVaw2TwbXj5pESMzfRKPvVQ61o

¹⁸ <https://aws.amazon.com/pt/what-is/data-science/>

integra as bases de dados curriculares, grupos de pesquisa e instituições em um único sistema de informações¹⁹;

- **Análise Econométrica:** Uso de ferramentas estatísticas, tais como regressão linear, para buscar correlação (inferir o impacto) entre duas ou mais variáveis através de um modelo matemático²⁰;
- **Quase-Experimento:** estudo empírico usado para estimar o impacto causal de uma intervenção na população-alvo sem atribuição aleatória. A pesquisa quase-experimental compartilha semelhanças com o desenho experimental tradicional ou ensaio controlado randomizado, mas carece especificamente do elemento de atribuição aleatória de tratamento ou controle²¹;
- **Diferença em Diferenças:** metodologia que consiste em aplicar uma dupla diferença, ou seja, busca comparar as mudanças no tempo na variável de interesse entre o grupo de tratamento e o grupo de controle. Para que essa metodologia seja válida não é preciso que os grupos de tratamento e controle sejam semelhantes entre si em suas características observáveis. Ao contrário do experimento aleatório, onde a condição de equilíbrio é essencial na medição do impacto, no método de diferenças em diferenças a condição necessária para avaliar o impacto é a chamada suposição de tendências paralelas. Isso pressupõe que, na ausência de tratamento, os dois grupos teriam seguido a mesma trajetória nos resultados de interesse²².

Assim, para a análise bibliométrica, uniu-se os termos “Impacto”, “Ciência”, mais os abaixo elencados, relativos aos objetivos desta pesquisa, encontrando-se o seguinte quantitativo de artigos revisados por pares no Portal de Periódicos da CAPES:

- “Análise Econométrica” ou “Econometria”: 25;
- “Data Mining”, “Ciência de Dados”: 20;

¹⁹ <https://galoa.com.br/blog/historia-do-curriculo-lattes-o-que-e-e-por-que-usar/>

²⁰ <https://pt.wikipedia.org/wiki/Econometria>

²¹

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjHjfXAk_D-AhUnBbkGHSt-DZQQFnoECAwQAQ&url=https%3A%2F%2Fwww.passeidireto.com%2Farquivo%2F103164267%2Fconsucao-da-pesquisa-cientifica&usg=AOvVaw2Lv21cjEM2d4J-fZXu9C0W

²²

https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiZwZ7Qk_D-AhUblRkGHUzxAicQFnoECAgQAQ&url=https%3A%2F%2Fwww.caf.com%2Fpt%2Fpresente%2Fnoticias%2F2020%2F02%2Fnota-metodologica-como-implementar-diferencas-em-diferencas-para-medir-o-impacto%2F&usg=AOvVaw2B1vUBGNce_Mv8J3ZuXj_P

nuvem acima indicada. Ainda assim, considera-se a revisão realizada suficientemente ampla para indicar que não foram encontrados trabalhos semelhantes a este na literatura. A listagem de artigos lidos e analisados (ao todo, 723 publicações) podem ser encontradas na pasta “Levantamento Bibliométrico” no GitHub já indicado.

6. METODOLOGIA

A metodologia está apresentada no artigo: Inferindo a Causalidade entre Financiamento em Ciências e a Formação de Recursos Humanos: um Modelo Econométrico para Medir o Impacto de Políticas Públicas por um Quase-Experimento.

O Artigo se encontra no Anexo I desta tese e apresenta um método para inferir a correlação entre uma política pública de investimento em ciência e vários indicadores presentes no Currículo Lattes do CNPq. Apresenta-se aqui um resumo do que é lá exposto.

Primeiramente, define-se “impacto” como “a medida do retorno social, cultural, ambiental e econômico de uma pesquisa financiada pelo setor público, sejam produtos ou ideias” e uma revisão bibliográfica sobre o assunto é realizada. São apresentadas as ferramentas para a análise de impacto e escolhidas aquelas quantitativas (Grupo 2): bibliometria, *data mining* e análise econométrica.

Enquanto Bibliometria pode ser definida como a “aplicação da matemática e métodos estatísticos para livros e outras mídias de comunicação”, a cientometria busca “os métodos quantitativos da pesquisa sobre o desenvolvimento da ciência como um processo informacional”²⁴. Assim, esta pesquisa vai além da ferramenta bibliométrica, uma vez que busca indicadores capazes de representar várias dimensões do impacto da pesquisa, e é proposto método de extração de indicadores presentes no Lattes para uso na medição de uma medida do impacto da pesquisa. Para tal, é proposta ferramenta de informática capaz de extrair esses indicadores, por meio de *Data Mining* no Currículo Lattes e de modo a representar diversas dimensões do impacto.

Finalmente, é estudada forma de implementação de uma análise econométrica. O modelo das diferenças em diferenças por meio de um quase-experimento é adotado. Esse é um método capaz de corrigir para diferenças observáveis e não observáveis e busca inferir a causalidade de um fenômeno. O modelo econométrico proposto é:

Equação 4 – Modelo Econométrico para o Cálculo do Impacto.

$$Indicador_{atual,i} = C + I * Indicador_{anterior,i} + P * Financiamento_i + \varepsilon_i$$

Fonte: Elaborado pelo autor.

Na Equação acima, “i” representa cada indivíduo, “C” representa uma constante e seria a progressão do indicador que independeria de qualquer financiamento e/ou produtividade anterior e constante no tempo. “I” representa a medida em que um conhecimento anterior facilita o conhecimento posterior e ε_i é o erro para cada indivíduo, de modo que P representa o

²⁴ <https://bsf.org.br/2015/03/24/diferenca-definicao-conceito-bibliometria-cientometria-infometria-altmetrics/>

impacto do *Financiamento*_{*i*}.

Como ferramenta matemática de cálculo do modelo econométrico proposto apresenta-se a regressão linear. Assim, o método proposto é exploratório, de abordagem quantitativa, escalável, transparente, que permite comparação entre avaliações diferentes e adequado a uso frequente e longitudinal. Traz medida quantitativa do impacto econômico, social, ambiental e cultural e permite o cálculo da eficácia do instrumento analisado. No entanto, não é flexível, não permite inferir causas ou demais efeitos, não é preditivo, explicativo, nem permite inferências além daquelas representadas pelos indicadores analisados. Portanto, é ineficaz para explicar os motivos do impacto medido ou para buscar impactos não representados pelos indicadores escolhidos e, portanto, inadequado como ferramenta exclusiva de escolha.

7. CONSIDERAÇÕES ADICIONAIS NESTA PESQUISA

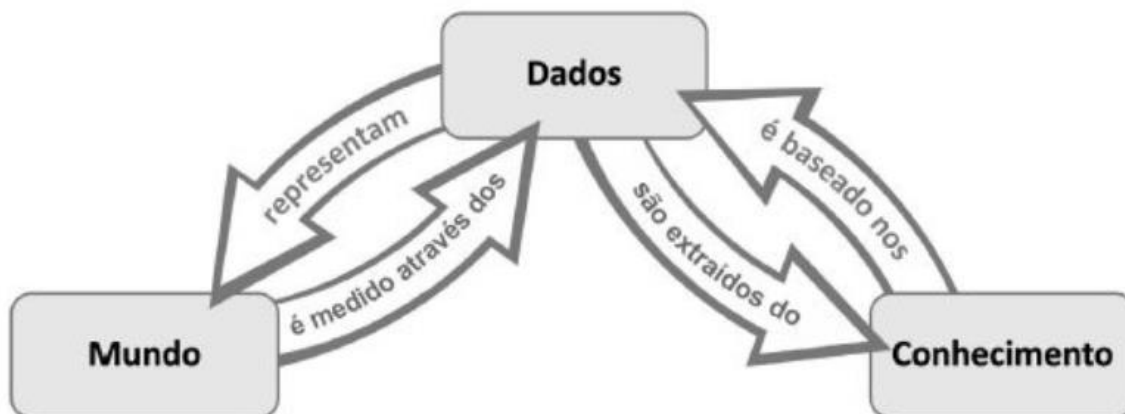
7.1. Ciência de Dados

Na bibliografia levantada, observa-se tendência a que o uso de análise de dados em *Big Data* não seja expressamente mencionado como ferramenta metodológica, mas subentendida no processo de coleta de dados. Dessa forma, embora tenha havido em muitas pesquisas a etapa de coleta, análise e tratamento dos dados, resultando em uma mineração passível de uso direto nas pesquisas encontradas, a tendência é a de que esse processo tenha sido realizado por um terceiro e apenas a origem dos dados apontada, ao invés de todo o processo detalhado na pesquisa. O encontrado está de acordo com o encontrado por (SALDANHA, 2022), que indica essa ser uma falha do *Big Data*, uma vez que os dados precisam de mineração e tratamento para serem utilizáveis na pesquisa, ocasionando insegurança e eventual imprecisão, uma vez que o processo implica interpretação e escolhas, muitas vezes transparentes aos pesquisadores.

Ocorre que, no caso desta pesquisa, não há terceiro responsável pela coleta e tratamento dos dados, toda a coleta e tratamento foi realizada mediante o Extrator Lattes e as linguagens SQL e Python, publicamente disponibilizadas, de modo que o uso da ciência de dados faz parte inerente da pesquisa, uma vez que qualquer erro cometido no processo se traduz em dados incorretos ou imprecisos. Assim, torna-se necessária uma revisão de literatura sobre o assunto e a indicação de como esta pesquisa coletou e tratou os dados.

Dados é tudo aquilo que é produzido pelo homem e busca representar a realidade. São exemplos de dados, uma anotação ou um retrato desenhado em um caderno de notas, que buscam descrever um ambiente, bem como a frequência cardíaca captada pelo *smartwatch*, uma receita média, um código CID, o caderno do aluno que anotou o conteúdo da aula, uma anotação de um biólogo sobre o diâmetro da folha, uma foto, um raio-x. Dessa forma, os exemplos são inúmeros, bem como os meios em que são coletados e a forma como são armazenados. Os dados podem ser transmitidos e replicados, permitindo um conhecimento mais amplo e preciso do que aquele adquirido por meio da simples observação pessoal, permitindo o conhecimento empírico (SALDANHA, 2022).

Figura 19 – A produção do conhecimento segundo a visão representativa dos dados

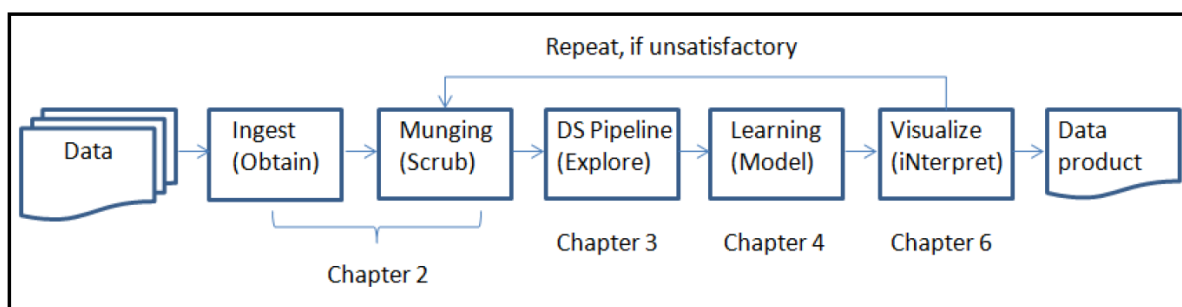


Fonte: © Sabina Leonelli, produzido por Michel Durinx. (SALDANHA, 2022)

Dessa forma, os dados não falam por si e sua validade depende originariamente de quem o produziu, a forma como o produziu, como o armazenou, sua completude, contemporaneidade, validade, aplicabilidade etc. Não são desprovidos de aspectos teóricos e subjetivos, precisam ser selecionados, tratados e interpretados, sendo que o conhecimento sobre o significado e a forma de captura dos dados pode ser essencial para sua compreensão e interpretação. Essa atividade é, portanto, altamente dependente de quem a faz, bem como a validade de seu resultado (SALDANHA, 2022).

A análise de dados inclui todas as técnicas e processos usados na extração de informações de dados brutos. Como os dados brutos geralmente não têm forma estruturada e são pouco informativos, a necessidade de organizar esses dados torna-se ainda mais importante. A Ciência de Dados envolve o processo de análise, tratamento (remoção de dados não utilizáveis, normalização, escolha, etc...) e exploração dos dados, criação de um modelo computacional para descrever os dados, aplicação do modelo e análise dos resultados, num ciclo que pode se repetir indefinidamente, conforma a necessidade.

Figura 20 – Processo de Ciência de Dados



Fonte: Python Data Science Essentials (ALBERTO BOSCHETTI; LUCA MASSARON, 2016)

A Ciência de Dados (Data Science) é considerada interdisciplinar na temática ou multidisciplinar na origem e, dentro da conjuntura de grandes volumes de dados (*Big Data*), há grande potencial tecnológico na pesquisa aplicada, com impacto em diferentes áreas do conhecimento e de setores de atuação. Nesta perspectiva, Conway (2010) criou o Diagrama de Venn para especificar as habilidades que concerne à área, em que a Ciência de Dados aparece no centro, em lugar de destaque, indicando a correlação com outras capacidades. Conway (2010) entende que apenas focar na Ciência de Dados não é o suficiente para a compreensão da área em sua totalidade (CONWAY, 2010 APUD RODRIGUES; DIAS, 2017).

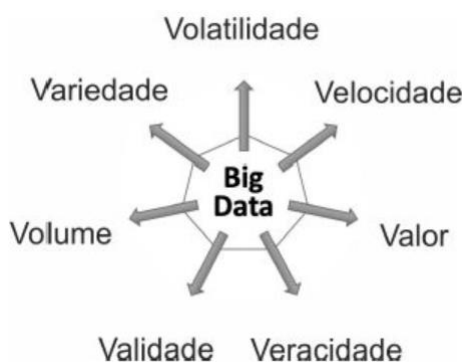
Figura 21 – Diagrama de Venn desenvolvido por Conway



Fonte: Estudos sobre visualização de dados científicos no contexto da Data Science e do Big Data (RODRIGUES; DIAS, 2017).

A ciência de dados está profundamente ligada ao Data Mining e ao *Big Data*, na medida em que busca estudar e extrair de um alto volume de dados informações úteis para o negócio em questão. Nesta pesquisa, considerou-se o Currículo Lattes como *Big Data*, isso porque atende aos requisitos necessários, que se caracteriza, de acordo com Saldanha (2022), pelos sete “Vs”:

Figura 22 – Os sete “Vs” do Big Data



Fonte: produzido por Michel Durinx (SALDANHA, 2022)

O Currículo Lattes possui Volatilidade, na medida em que se torna obsoleto rapidamente se não atualizado, Velocidade na medida em que é constantemente atualizado, Valor por buscar representar a vida e a formação dos pesquisadores brasileiros, Veracidade por se constituir de documento público e até penalização por inveracidades, Validade por ser preciso, pois espera-se que cada pesquisador complete o currículo da melhor forma possível, Volume, pois são mais de sete milhões de currículos e Variedade, pois não só pesquisadores que preenchem o currículo, mas a maior parte da população acadêmica do Brasil. O volume de dados da análise realizada chega a 1 Tb de dados, sendo que só o banco de dados dos dados já normatizados e tratados ocupa 268Gb de espaço em disco.

Saldanha (2022), em seu livro intitulado “A pesquisa científica na era do Big data: cinco maneiras que mostram como o Big data prejudica a ciência, e como podemos salvá-la”, mostra cinco perigos relacionados ao uso de Big Data: o uso de dados antigos; a insegurança e o uso de dados não confiáveis; parcialidade dos dados ou viés de seleção; corrupção (dados desonestos); e mal uso de dados sensíveis. Nesse sentido, esta pesquisa buscou observar as orientações de Saldanha (2022), buscando ser atual, confiável, imparcial, honesta e de modo a preservar os dados sensíveis.

Dada esta pequena introdução, o modo como os dados foram adquiridos e tratados será explicado nos artigos que seguem.

7.2. O Viés do Duplo Financiamento

Como será explicado, um dos campos presente no Currículo Lattes é se a pesquisa foi ou não financiada por uma agência pública de fomento. O duplo financiamento pode apresentar viés na análise, uma vez que o modelo econométrico adotado não o prevê. Por outro lado, o método das diferenças em diferenças já procura corrigir por fatores observados e não observados, como, por exemplo, por um financiamento externo. Dessa forma, mesmo presente eventual viés por duplo financiamento, esse seria, senão totalmente, ao menos parcialmente eliminado pelo método escolhido.

De qualquer forma, para evitar eventual viés por duplo financiamento e melhorar a precisão do impacto medido, da análise realizada com os doutorandos que entraram no ano de 2014 forem retirados todos os duplo-financiamentos da CAPES e do CNPq, ou seja, foram retirados da análise aqueles bolsistas que chegaram a receber bolsas de ambas as agências. Essa exclusão é necessária pois, em tese, não pode haver duplo financiamento concomitantemente entre essas duas agências, ou seja, o financiamento pela CAPES a doutorandos influencia

(exclui) o financiamento pelo CNPq e vice-versa (o que indica que o duplo financiamento ocorreu em momentos distintos e não ao mesmo tempo).

Observa-se que a CAPES não realiza financiamento direto a projetos de pesquisa, como o CNPq faz (exemplo, o Universal 2014 analisado). Dessa forma, não há exclusão a ser realizada por essa agência de fomento nesse estudo em particular.

Ainda, há forte probabilidade de que ao menos alguns doutorandos ou pesquisadores tenham recebido financiamento por outra agência de fomento, normalmente as agências estaduais (FAPESP, FAPERJ, etc.). No entanto, esses financiamentos, se presentes, constituem variáveis não observadas independentes e o método proposto busca já considerar essas no impacto medido. Assim, a pergunta que se deve fazer é: não tivesse o CNPq investido, esses pesquisadores ainda receberiam financiamento pelas outras agências de fomento? A resposta é, muito provavelmente, sim. Se isso for verdade, o investimento por outras agências pode ser interpretado como evento externo e independente. Se assim for considerado, o método das diferenças em diferenças que mede o impacto deve levar em conta também esse financiamento.

Explicando por outro lado, vemos que um dos impactos do Universal 2014 foi a formação de doutores. Pode-se argumentar que há essa formação somente porque os investimentos realizados foram destinados a professores doutores dentro de instituições com PPG. Ou seja, se não fosse o salário do professor e toda uma infraestrutura por trás, não haveria a formação de doutores. Isso é de fato verdade, mas a realidade é que foram os professores, doutores, dentro de PPGs que foram financiados e esse é um fato independente, uma variável externa não observada.

Dessa forma, o impacto medido do Universal 2014 indicou que esse financiamento teve um grande retorno em forma de benefício marginal em formação de doutores. Ou seja, um pequeno investimento a mais pelo CNPq, somado a todos os outros que já são investidos de qualquer forma, resulta num grande impacto em formação de doutores. Se o CNPq investir concedendo subsídio a projetos de pesquisa por meio do Universal, haverá formação de doutores. Se o CNPq não investir, não haverá. Assim, essa formação é um impacto do investimento, já corrigido para as variáveis externas não observadas, como a qualidade da instituição em que o pesquisador se insere, financiamentos externos, o salário de professor pago, a cultura de *publish or perish*, etc.

Assim, o modelo calcula o impacto do programa *ceteris paribus* (tudo o mais mantido constante), um benefício marginal ou, em outras palavras, o que o investimento realizado pelo CNPq, mantidos todos os outros investimentos e todas as outras variáveis constantes, traz a mais (ou a menos).

7.3. A dimensão da Qualidade e a falta de análise dessa dimensão neste estudo.

Uma das poucas dimensões do impacto do financiamento que o Currículo Lattes *per si* não abrange é a dimensão da qualidade. O Extrator Lattes não traz, junto ao currículo, qualquer indicador de qualidade da publicação, como o número de artigos ou a qualidade da revista publicada.

Indicador da qualidade da revista poderia ser buscado, fazendo-se a conexão entre o DOI do artigo, o ISSN da revista ou mesmo o nome da mesma e planilha fornecida, por exemplo, pela CAPES. No entanto, há diversos alertas internacionais quanto ao uso do fator de impacto da revista como medida de qualidade do artigo publicado.

Entre os muitos estudos encontrados no assunto, apresento a DORA²⁵ (The Declaration on Research Assessment), que apresenta a seguinte recomendação: “a necessidade de eliminar o uso de métricas baseadas em periódicos, como o fator de impacto, nas considerações de financiamento, nomeação e promoção” e “a necessidade de avaliar a pesquisa por seus próprios méritos, em vez de depender dos méritos do periódico no qual ela foi publicada” (BIOLOGY, 2012).

Um dos objetivos desta pesquisa é possibilitar a metodologia desenvolvida de modo escalável horizontalmente (entre políticas públicas) e verticalmente (no tempo), inclusive no futuro. Dessa forma, optou-se por não usar uma métrica de qualidade condenada mundialmente, muito embora ainda muito utilizada.

A métrica mais aceita como medida de qualidade de um artigo científico é sua quantidade de citações, muito embora haja diversas críticas a esse indicador como medida de qualidade (LINDSEY, 1989). O artigo citado, “Using citation counts as a measure of quality in science measuring what's measurable rather than what's valid”, é uma referência na matéria e ainda hoje citado e tido como atual.

O artigo de Lindsey indica que o número de citações, embora ainda seja o que melhor represente a qualidade da pesquisa, é sensível à popularidade do tema, é útil apenas para separar os extremos, não podendo ser adequadamente utilizado para classificar a qualidade de um artigo, não é sensível à dimensão ética e moral da ciência e ainda subestima a contribuição da ciência aplicada.

Assim, buscou-se um método automatizado capaz de trazer a quantidade de citações por artigo encontrado. Não foi encontrado método gratuito para essa análise de citações, de modo que a ausência dessa dimensão é uma das limitações deste estudo.

²⁵ <https://sfdora.org/read/read-the-declaration-portugues-brasileiro/>

8. O IMPACTO DE BOLSAS DE DOUTORADO

Com o uso da metodologia proposta, também foi calculado o Impacto da Concessão de Bolsas de Doutorado pelas duas agências de fomento federal brasileiras (a CAPES e o CNPq). A pesquisa resultou no artigo, que se encontra no Anexo II: O Impacto de Bolsas de Doutorado na Pós-Graduação Brasileira: uma Análise pelo Currículo Lattes do CNPq. Apresentamos aqui apenas um breve relato do estudo.

A população escolhida consistiu daqueles que iniciaram seu primeiro doutorado no ano de 2014 e o modelo econométrico adotado foi inicialmente:

Equação 5 – Medida de um Impacto por meio de Regressão multivariada categórica

$$Indicador_{2019} = \alpha + \beta Indicador_{2013} + \gamma CAPES + \delta CNPq$$

Em que $Indicador_{2013}$ e $Indicador_{2019}$ representam a soma de todas as entradas desse indicador até os anos de referência (2103 e 2019), CAPES é uma variável categórica igual a 1, caso o pesquisador tenha recebido bolsas dessa agência, ou 0 caso não tenha, sendo a mesma metodologia aplicada à variável CNPq, α representa um aumento constante do indicador no tempo, β o aumento temporal desse indicador em relação a seu estado inicial e aqueles com duplo financiamento foram retirados da pesquisa. Com esse modelo, chegou-se no seguinte impacto:

Tabela 3 – Impacto de Concessão de Bolsas de Doutorado pelo CNPq e pela CAPES.

Indicador	Nº CAPES	Impacto CAPES	Nº CNPq	Impacto CNPq	Impacto Total
Artigo Publicado	7833	10856	2297	4163	15.019
Trabalho em Eventos	7493	7084	2122	1844	8.927
Detalhamento do Trabalho	7476	6790	2118	1683	8.473
Novo Vínculo Empregatício ou Funcional	142	4951	57	2247	7.198
Apresentação de Trabalho Técnico	7653	3251	2183	0	3.251
Conclusão de Curso de Curta Duração	6770	1764	1851	671	2.435
Ministério de Curso de Curta Duração	4243	1807	1103	577	2.384
Produção de Trabalho Técnico	2675	2060	669	0	2.060
Nova Atividade de Ensino	404	1454	117	353	1.807
Prêmio ou Título Recebido	4521	586	1416	0	586
Participação em Simpósio	6098	535	1754	0	535
Participação em Congresso	7609	0	2210	531	531
Novo Projeto de Pesquisa	685	-176	186	428	253
Produção de Carta, Mapa ou Similar	117	229	26	0	229

Nova Atividade de Pesquisa e Desenvolvimento	1341	125	330	-6	119
Novo Estágio	60	1	11	7	8
Artigo Aceito para Publicação	1561	-192	421	-65	-257
Publicação de Capítulo de Livro	4432	0	1156	-312	-312
Outras Orientações Concluídas	4721	-6265	1233	-2372	-8.637
Participação em Banca de Graduação	5307	-6330	1325	-2856	-9.185

Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Como a informação sobre o montante financiado estava disponível apenas para aquelas bolsas financiadas pelo CNPq, o seguinte modelo econométrico foi aplicado a esse financiamento:

Equação 6 – Medida de Impacto por meio de Regressão multivariada categórica para o CNPq:

$$Indicador_{2019} = \alpha + \beta Indicador_{2013} + \theta Financiamento_{CNPq}$$

O Grupo de Controle foi escolhido pelo método de pareamento, em que se buscou o Currículo Lattes o mais próximo possível daqueles pertencentes ao grupo de tratamento. Com os dados, foi possível chegar na seguinte tabela de custo-efetividade:

Tabela 4 – Custo-Efetividade da concessão e Bolsas a Doutorandos 2014 por concessão ou não e por valor investido.

Indicador	Custo-Efetividade por Concessão ou não	Custo-Efetividade por Pareamento	Média
Participação em Banca de Graduação	-R\$ 101.514,21	-R\$ 88.553,63	-R\$ 95.033,92
Outras Orientações Concluídas	-R\$ 122.227,90	-R\$ 84.329,43	-R\$ 103.278,67
Artigo Publicado	R\$ 69.643,19	R\$ 150.688,46	R\$ 110.165,82
Novo Vínculo Empregatício ou Funcional	R\$ 129.027,41	R\$ 226.680,68	R\$ 177.854,04
Orientação em Andamento de Aperfeiçoamento ou Especialização	#N/D	-R\$ 133.667,40	-R\$ 133.667,40
Trabalho em Eventos	R\$ 157.225,92	#N/D	R\$ 157.225,92
Detalhamento do Trabalho	R\$ 172.266,54	#N/D	R\$ 172.266,54
Orientação em Andamento de Graduação	#N/D	-R\$ 280.391,29	-R\$ 280.391,29
Conclusão de Curso de Curta Duração	R\$ 432.078,37	#N/D	R\$ 432.078,37
Início de Curso de Curta Duração	R\$ 435.322,21	#N/D	R\$ 435.322,21
Ministério de Curso de Curta Duração	R\$ 502.468,96	#N/D	R\$ 502.468,96
Participação em Congresso	R\$ 545.997,34	R\$ 494.751,86	R\$ 520.374,60
Novo Projeto de Pesquisa	R\$ 677.393,90	#N/D	R\$ 677.393,90
Participação em Encontro	#N/D	R\$ 789.985,26	R\$ 789.985,26

Nova Atividade de Ensino	R\$ 821.316,12	#N/D	R\$ 821.316,12
Publicação de Capítulo de Livro	-R\$ 929.245,48	#N/D	-R\$ 929.245,48
Início de Residência Médica	#N/D	R\$ 1.213.073,60	R\$ 1.213.073,60

Fonte: Autoria própria (2023).

O estudo encontrou que mais de 50% dos doutorandos receberam bolsas e que houve impacto, tanto na CAPES quanto no CNPq, em diversos indicadores, incluindo vínculos empregatícios e número de artigos publicados. O estudo confirmou a importância da concessão de bolsas para a formação no nível de doutorado e um alto impacto positivo, mas também impacto negativo naquelas atividades menos especializadas, como participação em bancas de graduação.

O cálculo da custo-efetividade permite a comparação entre diversas políticas públicas, dando indicação de qual política possui uma efetividade melhor e, portanto, servindo como orientação. Alerta-se, no entanto, que sem uma análise mais detalhada a metodologia não consegue indicar qual a “melhor” política pública.

Em consideração final, os investimentos praticados, tanto pela CAPES quanto pelo CNPq, no geral, têm impacto positivo e pode-se dizer que proporcionaram, aproximadamente, a publicação de 15.019 artigos, 8.927 trabalhos em eventos, 7.198 novos vínculos, 2.060 trabalhos técnicos, 1.807 novas atividades de ensino, entre diversos outros, o que colabora com as metas do PNE e com o desenvolvimento socioeconômico do País.

9. O MODELO ECONOMETRICO E A REDUÇÃO DE DIMENSÕES PROPOSTAS

Um dos questionamentos levantados durante o desenvolvimento do modelo foi se esse não está simplificado, ou seja, se não há outros indicadores (variáveis independentes) que deveriam ser inclusos no modelo adotado:

Equação 7 – Regressão Linear entre índice atual, índice anterior e financiamento.

$$Indicador_{atual,i} = C + I * Indicador_{anterior,i} + P * Financiamento_i + \varepsilon_i$$

Fonte: Elaborado pelo autor.

O primeiro passo para a definição do modelo a ser adotado é a análise estatística exploratória, que busca definir quais variáveis independentes de fato afetam a variável dependente, qual o modelo que faz mais sentido teórico e quais os seus pressupostos.

Para a análise exploratória, realizou-se um estudo *stepwise*, em que todos os 142 indicadores (140 do Lattes, a constante mais a quantia de pagamentos realizados) foram utilizados para modelar o indicador artigos publicados até o ano de 2019. A metodologia *stepwise* é aquela em que a programação elimina o indicador menos significativo, um a um, até que todos sejam estatisticamente significativos ao parâmetro estabelecido (aqui, 0,01 ou 99%). Os dados nesta análise exploratória foram retirados de doutorandos 2014, sendo o grupo de tratamento aquele financiado pelo CNPq (e não pela CAPES) e o grupo de controle eleito por pareamento (ver o O Impacto de Bolsas de Doutorado).

Dessa forma, realizando o *stepwise*, restaram 19 indicadores estatisticamente significativos, de acordo com a tabela abaixo²⁶:

Regressão 1 – Regressão StepWise correlacionando o nº de artigos publicados em 2019 com o estado do Currículo Lattes em 2014 e financiamento pelo CNPq.

Model:	OLS	Adj. R-squared:	0.675
Dependent Variable:	qty_2019	AIC:	28793.4891
Date:	2023-05-23 17:18	BIC:	28916.6821
No. Observations:	4836	Log-Likelihood:	-14378.
Df Model:	18	F-statistic:	560.0
Df Residuals:	4817	Prob (F-statistic):	0.00
R-squared:	0.677	Scale:	22.470

²⁶ Código-Fonte: 21. Verificando Correlação entre Patente e Doutorado.ipynb

	Coef.	Std.Err.	t	P> t 	[0.025	0.975]
Intercept	1.7003	0.1180	14.4140	0.0000	1.4690	1.9315
porcent_pagos	1069.5866	249.5781	4.2856	0.0000	580.2996	1558.8736
detalhamento_do_trabalho	0.0855	0.0070	12.2485	0.0000	0.0718	0.0992
artigo_publicado	1.6276	0.0235	69.3329	0.0000	1.5816	1.6737
fim_de_atividade_de_pesquisa_e_desenvolvimento	-0.9293	0.1931	-4.8135	0.0000	-1.3077	-0.5508
inicio_de_atividade_de_ensino	-0.2907	0.0943	-3.0830	0.0021	-0.4756	-0.1059
inicio_de_atividade_de_pesquisa_e_desenvolvimento	0.6142	0.1323	4.6431	0.0000	0.3549	0.8736
participacao_em_seminario	-0.0913	0.0217	-4.2134	0.0000	-0.1338	-0.0488
participacao_em_simposio	0.1275	0.0301	4.2389	0.0000	0.0685	0.1865
conclusao_de_especializacao	-0.3062	0.0913	-3.3516	0.0008	-0.4852	-0.1271
fim_de_atividade_de_ensino	0.3036	0.1024	2.9638	0.0031	0.1028	0.5044
fim_de_projeto_de_pesquisa	-0.8730	0.0708	-12.3251	0.0000	-1.0118	-0.7341
inicio_de_projeto_de_pesquisa	0.8243	0.0563	14.6447	0.0000	0.7139	0.9346
participacao_em_encontro	-0.0885	0.0217	-4.0682	0.0000	-0.1311	-0.0458
artigo_aceito_para_publicacao	-0.8989	0.2480	-3.6243	0.0003	-1.3852	-0.4127
producao_demais_trabalhos	-0.2727	0.0809	-3.3712	0.0008	-0.4313	-0.1141
inicio_de_atividade_de_conselho_comissao_ou_consultoria	-0.2319	0.0651	-3.5596	0.0004	-0.3596	-0.1042
conclusao_de_residencia_medica	-0.7511	0.2550	-2.9462	0.0032	-1.2509	-0.2513
manutencao_de_obra_artistica	-0.2175	0.0614	-3.5436	0.0004	-0.3378	-0.0972
Omnibus:	4850.969	Durbin-Watson:	1.991			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	926863.638			
Skew:	4.441	Prob(JB):	0.000			
Kurtosis:	70.238	Condition No.:	59942			

O grau de significância a 99% (0.01 para a hipótese nula) foi escolhido porque, como há mais de uma centena de indicadores, mesmo com esse grau de significância ainda há probabilidade estatística de termos um ou dois indicadores presentes que não estão efetivamente correlacionados à quantidade de artigos publicados até o ano de 2019 (indicador presente devido ao acaso).

Ao analisarmos os indicadores acima no método exploratório *stepwise*, verificamos que a maior parte das correlações faz sentido, mas que não devem entrar no modelo adotado. Como exemplo, participações em simpósios e seminários estão normalmente correlacionados à publicação de um artigo e a uma atividade de pesquisa e desenvolvimento. Ou seja, há correlação entre as variáveis independentes, o que justifica a exclusão delas do modelo.

No entanto, observa-se que, na regressão acima realizada, $R^2_{ajustado}=0.675$. Eliminando todas as variáveis independentes, salvo a constante, o indicador no ano de 2014 e os pagamentos realizados, que é o modelo utilizado neste estudo, eliminamos várias correlações entre essas variáveis (co-dependências das variáveis independentes) e temos o seguinte resultado:

Regressão 2 – Regressão correlacionando o nº de artigos publicados em 2019 com o nº de artigos publicados em 2014 e financiamento pelo CNPq.

Results: Ordinary least squares

```

=====
Model: OLS                               Adj. R-squared: 0.628
Dependent Variable: qty_2019             AIC: 29441.6545
Date: 2023-05-23 17:27                  BIC: 29461.1060
No. Observations: 4836                  Log-Likelihood: -14718.
Df Model: 2                             F-statistic: 4076.
Df Residuals: 4833 P                    rob (F-statistic): 0.00
R-squared: 0.628                         Scale: 25.778
-----

```

	Coef.	Std.Err.	t	P> t	[0.025	0.975]
Intercept	1.8973	0.0976	19.4358	0.0000	1.7059	2.0886
artigo_publicado	1.8635	0.0208	89.4436	0.0000	1.8227	1.9044
porcent_pagos	1944.5	258.7591	7.5150	0.0000	1437.2	2451.8587

```

-----
Omnibus: 4543.163          Durbin-Watson: 2.005
Prob(Omnibus): 0.000      Jarque-Bera (JB): 585986.396
Skew: 4.086               Prob(JB): 0.000
Kurtosis: 56.304         Condition No.: 14720
=====
* The condition number is large (1e+04). This might indicate
strong multicollinearity or other numerical problems.

```

Ou seja, o $R^2_{ajustado}$ caiu de 0.675 para 0.628, uma perda de apenas 0,047 (menos de 5% de previsibilidade), ao benefício de eliminarmos 16 indicadores da regressão (perda de aproximadamente 0,3% de previsão por indicador retirado). A simplificação realizada no modelo indica que há pouca dependência entre o indicador de 2019 e outros indicadores de 2014, embora exista a correlação estatisticamente significativa. Ou seja, pressupõe que a quantidade de artigos publicados até o ano de 2019 pouco depende, por exemplo, de quantos simpósios o doutorando participou até o ano de 2014. Assim, a suposição de que o modelo adotado é o melhor é de fato forte, mas, como já indicado, a perda de apenas 0,3% por indicador na previsão realizada indica que o pressuposto pode ser acatado.

Além disso, observa-se que não propomos apenas um modelo, mas 140 diferentes modelos que, em conjunto, calcularão eventual impacto. Ou seja, cada um dos 140 indicadores será considerado como possível impacto (variável dependente em t_1 e também como variável independente em t_0). Dessa forma, se não fizermos a redução de dimensões proposta, cada indicador em t_0 será considerado múltiplas vezes no modelo adotado, podendo provocar subvalorização do impacto medido.

Observa-se que o modelo adotado é simples e direto. Assim, é necessário trazer algumas considerações:

- Teoricamente, espera-se que a relação de um indicador, como o número de artigos publicados, no ano de 2019 seja dependente principalmente de quantos indicadores tinha no ano de 2014 e do financiamento realizado, além de constante e do erro inerente. Dessa forma, o modelo pressupõe *ceteris paribus* nos demais indicadores e condições.
- Por ser adotada a metodologia de diferenças em diferenças com grupo de controle artificialmente selecionado para redução do viés, pressupõe-se que o impacto medido não depende de características observáveis ou não observáveis desses grupos, mas da manutenção e equivalência da inclinação das retas no tempo dos dois grupos. Ou seja, mesmo que o pressuposto anterior não seja verdadeiro, a metodologia indicada permite inferir o impacto com a utilização da redução de dimensões indicada.
- Não se trata de regressão temporal, pois apenas dois anos estão sendo estudados. Dessa forma, o modelo não pressupõe ou assume eventual linearidade da curva, mas apenas que a eventual tendência não linear possa ser traduzida pela mesma média da inclinação linear da regressão realizada. Ou seja, que a curva no tempo do grupo de tratamento seja similar (mesmo que deslocada acima ou abaixo) da curva no tempo do grupo de controle.
- Por não se tratar de regressão temporal, nada pode ser inferido sobre a forma da curva. Outro estudo seria necessário para inferir esse formato.
- O modelo é usado para cálculo estatístico das diferenças em diferenças e mede, portanto, apenas uma média estatística do impacto. Não é, portanto, preditivo. Em outras palavras, não se pode usar o modelo para buscar inferir quantos artigos seriam publicados em 2018 ou em 2020, ou mesmo inferir quantos artigos um determinado pesquisador teria dado o estado do seu currículo em 2014.

Conclui-se que, como não se busca um modelo preditivo, mas o cálculo do impacto, o modelo adotado deve ser o mais simples possível, inclusive porque a inclusão de variáveis independentes forçosamente reduzirá o impacto medido e aumentará o R^2 , mesmo que não faça sentido teórico, causando uma subvalorização desse impacto. Como “prova”, o impacto medido pelas diferenças em diferenças diretamente aplicada ao total dos indicadores (sem o uso da regressão linear) é maior que o impacto calculado por meio da regressão linear, o que já demonstra que o impacto medido pela regressão linear já se encontra subvalorizado. Assim, acrescentar variáveis independentes melhoraria a previsibilidade do modelo em 5%, mas essa não é a intenção.

Há de se lembrar também que, mesmo utilizando-se uma significância de 0,01, como há mais de 100 indicadores, usar todos como variáveis independentes forçosamente significaria em alta probabilidade de adicionarmos um ou dois puramente devido a uma coincidência estatística. Dessa forma, a inclusão de mais variáveis independentes somente faria sentido em um modelo preditivo e não faz sentido teórico nesta pesquisa.

Com as considerações acima realizadas, uma sugestão de próximo estudo a ser realizado é ampla análise de redução de dimensões, de forma a agrupar os 140 indicadores (preferencialmente mais alguns relativos à qualidade da pesquisa) em um certo número reduzido de dimensões capazes de evitar em parte a correlação entre esses indicadores e facilitando a adoção de outro modelo, este com a finalidade preditiva ou de fornecer suporte para eventual avaliação do pesquisador.

10. O IMPACTO DO UNIVERSAL 2014

Com o uso da metodologia proposta, também foi calculado o Impacto e a Custo-Efetividade do Edital Universal 2014. A pesquisa resultou no artigo O Impacto de um investimento público em ciência na formação de recursos humanos em nível de mestrado e doutorado: o caso do Universal 2014.

O Artigo se encontra no Anexo III desta tese e aplica o método apresentado para avaliar um investimento em pesquisa científica (Universal 2014) realizado pelo governo brasileiro por meio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). O modelo econométrico utilizado foi:

$$Indicador_{atual} = C + I * Indicador_{anterior} + P * Financiamento$$

Equação 8 - Regressão Linear entre índice atual, índice anterior e financiamento.

Indicador representa a variável estudada, como por exemplo, orientação de doutorado. *C* é uma constante que reflete o estado inicial e também inclui o erro, *I* dá a correlação de fatores externos não temporais, entre o *Indicador_{anterior}* e o *Indicador_{atual}* e *P* mede a correlação do financiamento com o *Indicador_{atual}*.

Com o modelo econométrico e a metodologia proposta, chegou-se ao seguinte impacto do Universal 2014, sendo que foi financiado o total de R\$ 194.101.318,26 em 5.222 projetos:

Tabela 5 – Impacto da Chamada Universal

Indicador	Impacto
Artigo Publicado	6071
Novos Vínculos Empregatícios ou Funcionais	1236
Participação em Banca de Doutorado	1119
Participação em Banca de Mestrado	899
Orientação Concluída de Doutorado	604
Trabalhos em Eventos	562
Produção de Trabalho Técnico	462
Desenvolvimento de Material Didático ou Instrucional	413
Publicação de Capítulo de Livro	321
Prêmio ou Título Recebido	316
Apresentação de Trabalho Técnico	309
Orientação Concluída de Mestrado	294
Participação em Congresso	202
Orientação Concluída de Pós-Doutorado	193
Participação em Banca de Exame de Qualificação	162

Texto em Jornal ou Revista	103
Produção de Software	96
Orientação em Andamento de Doutorado	91
Livro Publicado ou Organizado	78
Participação em Simpósio	73
Outras Participações de Bancas Julgadoras	71
Produção Bibliográfica	55
Outras Participações em Congressos	49
Patente Registrada	44
Editoração	44
Organização de Evento	38
Participação em Encontro	36
Participação em Oficina	32
Ministério de Curso de Curta Duração	32
Novas Extensões Universitárias	29
Conclusão de Extensão Universitária	26
Produção de Artes Cênicas	25
Outra produção Técnica	25
Publicação de Prefácio ou Posfácio	25
Participação em Exposição	21
Programa de Rádio ou TV	21
Participação em Banca Julgadora para Professor Titular	20
Treinamento Ministrado	14
Produção Técnica - Processos ou Técnicas	11
Produção Artística ou Cultural	11
Participação em Seminário	11
Orientação em Andamento de Pós-Doutorado	9
Tradução	8
Orientação em Andamento de Aperfeiçoamento ou Especialização	7
Novas Atividades Técnico/Científicas	6
Produção de Carta, Mapa ou Similar	4
Conclusão de Outra Formação Complementar	3
Participação em Banca Julgadora para Concurso Público	3
Produção de Artes Visuais	3
Produção - Demais Trabalhos	3
Novas Atividades de Conselho, Comissão ou Consultoria	1
Participação em Banca Julgadora para Avaliação de Cursos	1

Desenvolvimento de Desenho Industrial	-2
Conclusão de Especialização	-2
Mídia Social, website ou Blog	-3
Produção Técnica - Produto Tecnológico	-3
Participação em Banca Julgadora para Livre Docência	-5
Novos Estágios	-7
Novas Atividades de Pesquisa e Desenvolvimento	-8
Orientação em Andamento de Graduação	-10
Orientação em Andamento de Iniciação Científica	-11
Artigo Aceito para Publicação	-18
Novas Atividades de Ensino	-23
Participação em Direção ou Administração	-23
Orientação em Andamento de Mestrado	-27
Outras Orientações em Andamento	-28
Novas Atividades de Serviço Técnico-Especializado	-53
Conclusão de Pós-Doutorado	-54
Conclusão de Aperfeiçoamento	-67
Conclusão de Curso de Curta Duração	-82
Novos Projetos de Pesquisa	-155
Outras Orientações Concluídas	-331
Participação em Banca de Graduação	-716

Fonte: Produção própria.

Dessa forma, verificou-se alto impacto na formação de recursos humanos, superando em muitas ordens o financiamento realizado, devendo, portanto, ser uma Política Pública não só a ser mantida, mas incentivada e expandida. Além disso, uma vez que se tem tanto o impacto do investimento realizado, como também o montante investido, pode-se chegar em tabela que representa a custo-efetividade, permitindo a comparação entre diversas políticas públicas. Novamente, alerta-se que, apesar de permitir a referida comparação, sem uma análise mais detalhada, a metodologia não consegue indicar qual a “melhor” política pública.

11. PARTE IV - CONCLUSÃO

Este estudo realizou uma análise de políticas públicas de investimento em ciência. Foram analisadas questões de equidade e eficácia, incluindo o custo administrativo para se ter uma medida de custo-efetividade. Além disso, o incentivo ao esforço máximo e a permanência foram indiretamente medidos por meio do cálculo do impacto em diversos indicadores retirados do Currículo Lattes.

Na análise de equidade realizada, este estudo verificou um tipo de presença de misoginia num órgão público nunca antes estudado: o menor financiamento médio a mulheres em projetos de pesquisa, e a menor presença em posições de poder, ajudando a firmar a teoria de que mulheres encontram tanto um teto de vidro quanto um labirinto de cristal, enfrentando barreiras invisíveis que dificultam seu progresso na carreira científica.

Além disso, este estudo desenvolveu uma nova forma de análise de impacto em políticas públicas de financiamento em ciência, através de um modelo econométrico baseado no método das diferenças em diferenças e por meio da retirada de indicadores pelo uso da cientometria e ferramentas de *Big Data* no Currículo Lattes.

Finalmente, para testar a ferramenta desenvolvida, foi realizada avaliações em duas distintas políticas públicas: a de concessão de bolsas de doutorado, tanto pela CAPES, quanto pelo CNPq; bem como o investimento direto a projetos de pesquisa, em todas as áreas do conhecimento, por meio da análise de impacto do Universal 2014. O impacto de ambas as políticas foi calculado e uma análise comparativa será realizada adiante.

11.1. Conclusões sobre a Análise de Equidade – Misoginia no CNPq

Um estudo sobre o número de concessões realizadas pelo CNPq ao longo dos anos demonstrou que as mulheres têm apresentado aumento de sua presença constantemente, ao ponto de ultrapassar a presença masculina e atingir 53% das concessões. No entanto, um estudo mais aprofundado dos financiamentos concedidos demonstram que o CNPq está longe de atingir um padrão de equidade. Muito ao contrário, há forte indicador de presença de misoginia no órgão, uma vez que, conforme cai a publicidade dada ao ato e conforme aumenta o valor ou o status da concessão, menor é a presença feminina.

Se separarmos todas as concessões do órgão por faixa de valores e por ano, observa-se que a presença feminina cai na medida em que a faixa de valor aumenta e na medida em que nos aproximamos do ano de 2019, tanto no número de concessões quanto no valor financiado.

Observa-se queda maior no valor financiado que no número bruto de concessões. Assim, a desigualdade se acentuou ao longo dos anos e até o ano de 2019.

Como o valor financiado é um dado pouco divulgado e como temos que, quanto maior o valor da concessão menor é a presença feminina, observa-se que a desigualdade se acentua conforme os dados vão ficando menos visíveis à sociedade. Esse é um forte indicador de misoginia, de que a cultura da discriminação sexual permeia o CNPq e de que, consciente ou inconscientemente, essa discriminação busca se ocultar dos olhos da sociedade.

No entanto, nos anos de 2020 e 2021, houve tendência para a aproximação da equidade. Observa-se, no entanto, que esses foram os anos em que a Pandemia COVID-19 se desenvolveu e, portanto, nenhuma conclusão pode ser tirada do fato. O estudo deve, portanto, continuar nos anos seguintes para se inferir como a desigualdade feminina se desenvolverá.

Ainda, ao analisarmos o valor médio que cada mulher recebe por concessão, comparado ao valor médio recebido pelo homem, temos que mulheres recebem menos que homens, por projeto, principalmente em projetos cujos valores da concessão ultrapassam trezentos mil reais. Dessa forma, encontra-se evidência do teto de vidro no CNPq, em que posições de responsabilidade ou de liderança são dificultadas às pessoas do sexo feminino.

Observa-se que as mulheres só estão sub-representadas nas áreas de Ciências Exatas e da Terra e Engenharias, ultrapassando os homens, em termos de presença (número de concessões) em todas as outras áreas do conhecimento. Ainda assim, recebem menos, em média e por projeto, que homens em todas as áreas do conhecimento, exceto Engenharias (mas, de novo, estão em menor número aqui). Dessa forma, verifica-se que a misoginia se estende a todas as áreas do conhecimento.

Finalmente, observamos certa equidade em bolsas, com uma leve sub-representação em bolsas de graduação. No entanto, verificamos forte sub-representação em bolsas de Produtividade em Pesquisa (bolsas PQ), bem como em Apoio a Projetos de Pesquisa. Isso condiz com uma misoginia capaz de criar barreiras a que mulheres atinjam posição de controle ou *status*.

Dessa forma, se unirmos todos os dados, encontramos que, embora o CNPq conceda mais a mulheres do que a homens, em números brutos, há misoginia presente em todas as áreas do conhecimento, favorecendo os homens quando se trata de financiamento de mais alto valor ou posições de poder, ou concedendo um maior financiamento a homens quando se analisa projetos de pesquisa, fazendo com que uma aparente equidade esconda presença misógina.

Um estudo que se seguirá ao aqui apresentado será análises de Cox. A regressão de Cox busca modelo preditivo quando há possíveis várias variáveis independentes em distintos tempos. Assim, o modelo permite covariáveis dependentes do tempo e permite, em teoria, produzir função de sobrevivência capaz de prever a probabilidade da concessão, a depender do sexo.

Outro estudo que deverá ser realizado é a confirmação de quais dificuldades as mulheres encontram na carreira científica. Relembrando que, em entrevistas com três pesquisadoras, encontrou-se que mulheres são menos chamadas para colaborações científicas, mesmo quando estão fisicamente mais próximas e mais acessíveis, precisando usar de linguagem masculina em seus trabalhos, tendo jornada extra de trabalho como cuidadoras das famílias, sendo desfavorecidas na hora de assumir cargos de poder e sofrendo assédio constante. No entanto, o encontrado foi por meio de entrevistas informais e precisa ser confirmado antes de ser considerado verdadeiro.

11.2. Conclusões sobre o Impacto de Políticas Públicas em Ciências

Este estudo desenvolveu um modelo econométrico baseado no método das diferenças em diferenças para inferir a causalidade entre indicadores retirados do Currículo Lattes e assim obter uma medida de impacto de uma Política Pública de financiamento em ciência. O modelo foi então aplicado às duas grandes políticas públicas do CNPq, calculando o impacto na concessão de bolsas de doutorado, bem como na Chamada Pública Universal MCTI/CNPq nº 014/2014. As análises realizadas permitem as seguintes comparações de impacto entre as políticas públicas:

- Comparação entre o impacto da concessão de bolsas de doutorado entre a CAPES e o CNPq;
- Comparação entre o impacto da concessão de bolsas de doutorado e no investimento no Universal 2014, ambos realizados pelo CNPq.

11.2.1. Comparação entre o impacto da concessão de bolsas de doutorado entre a CAPES e o CNPq;

Comparando o Impacto dos investimentos realizados pela CAPES e pelo CNPq, encontramos a seguinte relação:

Tabela 6 – Impacto de Concessão de Bolsas de Doutorado pelo CNPq e pela CAPES.

Indicador	Nº CAPES	Impacto CAPES	Nº CNPq	Impacto o CNPq	Impacto Total
Artigo Publicado	7833	10856	2297	4163	15.019
Trabalho em Eventos	7493	7084	2122	1844	8.927
Detalhamento do Trabalho	7476	6790	2118	1683	8.473
Novo Vínculo Empregatício ou Funcional	142	4951	57	2247	7.198
Apresentação de Trabalho Técnico	7653	3251	2183	0	3.251
Conclusão de Curso de Curta Duração	6770	1764	1851	671	2.435
Ministério de Curso de Curta Duração	4243	1807	1103	577	2.384
Produção de Trabalho Técnico	2675	2060	669	0	2.060
Nova Atividade de Ensino	404	1454	117	353	1.807
Prêmio ou Título Recebido	4521	586	1416	0	586
Participação em Simpósio	6098	535	1754	0	535
Participação em Congresso	7609	0	2210	531	531
Novo Projeto de Pesquisa	685	-176	186	428	253
Produção de Carta, Mapa ou Similar	117	229	26	0	229
Nova Atividade de Pesquisa e Desenvolvimento	1341	125	330	-6	119
Novo Estágio	60	1	11	7	8
Artigo Aceito para Publicação	1561	-192	421	-65	-257
Publicação de Capítulo de Livro	4432	0	1156	-312	-312
Outras Orientações Concluídas	4721	-6265	1233	-2372	-8.637
Participação em Banca de Graduação	5307	-6330	1325	-2856	-9.185

Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Em negrito encontra-se a agência com a melhor custo-efetividade, ou seja, dado certo financiamento, a agência que conseguiu o maior impacto ao menor custo (no caso de impactos negativos, a melhor custo-efetividade é o impacto de menor módulo). Observa-se que as agências possuem impactos semelhantes entre si, sendo que o CNPq se sobressai em artigos publicados, cursos de curta duração e congressos, e a CAPES se sobressaindo em eventos e trabalhos técnicos.

Podemos afirmar que a CAPES teve forte presença, com pouca presença do CNPq, nos seguintes indicadores:

- Apresentação e Produção de Trabalho Técnico;
- Prêmio ou Título Recebido;
- Participação em Simpósio;
- Produção de Carta, Mapa ou Similar; e
- Nova Atividade de Pesquisa e Desenvolvimento.

Da mesma forma, o CNPq teve forte presença, com pouca presença pela CAPES, nos seguintes indicadores:

- Participação em Congresso
- Novo Projeto de Pesquisa

Assim, observa-se que os impactos são similares, mas que a CAPES apresenta um impacto um pouco mais amplo e menos direcionado a um projeto de pesquisa científica, enquanto o impacto das bolsas concedidas pelo CNPq se concentra em artigos, congressos e projetos de pesquisa. De qualquer forma, os indicadores não são capazes de indicar, *per si*, quais das duas seria a melhor política de investimento à ciência, como esperado, uma vez que possuem regras e objetivos semelhantes.

11.2.2. Comparação entre o impacto da concessão de bolsas de doutorado e no investimento no Universal 2014, ambos realizados pelo CNPq.

Uma vez que a custo-efetividade de duas distintas políticas públicas foram calculadas pela mesma metodologia, é possível a comparação entre essas políticas a fim de subsidiar o poder público sobre a melhor forma de aplicar os recursos nessas políticas públicas. Isso é feito na tabela abaixo.

Tabela 7 – Comparação da Custo-Efetividade no ano de 2014: Concessão de Bolsas de Doutorado e Universal.

Indicador	Custo-Efetividade Bolsas de Doutorado	Custo-Efetividade Universal 2014	Impacto Bolsas de Doutorado a cada 1 milhão investidos	Impacto Universal 2014 a cada 1 milhão investidos.
Artigo Publicado	110.165,82	31.971,89	9	31
Novos Vínculos Empregatícios ou Funcionais	#N/D	157.039,90	0	6
Participação em Banca de Doutorado	#N/D	173.459,62	0	6
Participação em Banca de Mestrado	#N/D	215.908,03	0	5
Orientação Concluída de Doutorado	#N/D	321.359,80	0	3
Participação em Congresso	520.374,60	960.897,62	2	1
Trabalhos em Eventos	#N/D	345.376,01	0	3

Produção de Trabalho Técnico	#N/D	420.132,72	0	2
Ministério de Curso de Curta Duração	502.468,96	6.065.666,20	2	0
Desenvolvimento de Material Didático ou Instrucional	#N/D	469.978,98	0	2
Conclusão de Curso de Curta Duração	432.078,37	-2.367.089,25	2	0
Prêmio ou Título Recebido	#N/D	614.244,68	0	2
Apresentação de Trabalho Técnico	#N/D	628.159,61	0	2
Orientação Concluída de Mestrado	#N/D	660.208,57	0	2
Participação em Encontro	789.985,26	5.391.703,29	1	0
Orientação Concluída de Pós-Doutorado	#N/D	1.005.706,31	0	1
Participação em Banca de Exame de Qualificação	#N/D	1.198.156,29	0	1
Início de Residência Médica	1.213.073,60	#N/D	1	0
Publicação de Capítulo de Livro	-929.245,48	604.677,00	-1	2
Texto em Jornal ou Revista	#N/D	1.884.478,82	0	1

Fonte: Autoria Própria

Podemos observar que, no geral, a custo-efetividade do Universal 2014 é melhor que o da concessão de bolsas de doutorado em todos os indicadores, exceto os de Curso de Curta Duração ou de Participação em Encontro, dois indicadores de pouca relevância científica e tecnológica.

Ainda, a cada um milhão investidos, há três novos doutores formados *ceteris paribus*, ou seja, mantidas constantes todas as outras variáveis observáveis ou não observáveis (benefício marginal). Uma bolsa de doutorado custa ao CNPq R\$ 148.800,00, de modo que um milhão investidos nessas bolsas daria, em tese, retorno de seis doutores. Ou seja, o investimento no Universal 2014 causa um impacto, além de todos os outros medidas, equivalente a mais meia bolsa de doutorado.

A conclusão a que se chega é que, pelos indicadores apresentados, é melhor investir no Universal do que na concessão de bolsas de doutorado. O método proposto não pode, contudo, indicar os motivos por trás da melhor custo-efetividade do Universal. A hipótese é a

de que a burocracia brasileira limita severamente o estudo científico. Como exemplo, é difícil para o pesquisador comprar os tubos de ensaio, reagentes ou o computador de que precisa. Assim, um pequeno investimento que permita gastos na ciência, livre de burocracia, se traduz em grande motivação (incentivo ao esforço máximo) e em alto impacto científico. Estudos de caso poderiam confirmar a hipótese levantada.

11.2.3. Terceira missão da universidade, o impacto social da ciência.

Observou-se pouco impacto social nos investimentos realizados. Em especial, eventuais impactos nos seguintes indicadores foram pequenos ou nulos:

- Apresentação de Obra Artística
- Apresentação em Rádio ou TV
- Composição Musical
- Desenvolvimento de Material Didático ou Instrucional
- Desenvolvimento de Topografia de Circuito Integrado
- Editoração
- Nova Participação em Direção ou Administração
- Livro Publicado ou Organizado
- Manutenção de Obra Artística
- Maquete
- Mídia Social, web site ou Blog
- Música
- Obra de Artes Visuais
- Participação em Exposição
- Participação em Feira
- Participação em Oficina
- Participação em Olimpíada
- Participação em Seminário
- Participação em Simpósio
- Patente Registrada
- Prêmio ou Título Recebido
- Produção de Obra Musical
- Produção Artística ou Cultural
- Produção de Artes Cênicas

- Produção de Artes Visuais
- Produção de Carta, Mapa ou Similar
- Produção de Software
- Produção de Sonoplastia
- Programa de Rádio ou TV
- Publicação de Capítulo de Livro
- Publicação de Partitura Musical
- Solicitação de Cultivar protegida
- Solicitação de Cultivar Registrada
- Texto em Jornal ou Revista

Observa-se, portanto, que o impacto medido se concentra na área científica, mas pouco se espalha para a terceira missão do estado, para o impacto social da ciência. Aqui, não se quer dizer que o impacto social é nulo, mas apenas que o impacto direto e nos indicadores acima listados é pequeno ou nulo.

O fato provavelmente decorre da cultura científica brasileira, que movimenta os pesquisadores a publicar artigos, preferencialmente artigos de qualidade, tendo todos os outros indicadores importância relativamente menor.

Esse fato é algo que deve ser profundamente estudado, pois, observa-se tendência contrária da indicada pela Tripla Hélice. Ou seja, o estado não está promovendo integração entre pesquisadores, indústria e comércio, mas apenas formando pesquisadores mais especializados e capacitados e avançando na fronteira do conhecimento. Embora esse avanço seja de fato importante, sem a absorção pela indústria e pelo comércio o impacto da ciência produzida é reduzido.

11.3. Considerações Sobre o Cálculo do Impacto

Como qualquer indicador, há de se ter certo cuidado ao interpretar impacto negativo. Por exemplo, um dos impactos negativos encontrados em ambos os estudos realizados foi o indicador “Artigo Aceito para Publicação”. No entanto, observa-se forte impacto no indicador “Artigo Publicado”. Se analisarmos ambos os indicadores, em conjunto, podemos observar que há menos artigos aceitos para publicação, mas mais artigos publicados, ou seja, poder-se-ia dizer que o tempo de publicação dos artigos diminuiu. Assim, um impacto que é aparentemente negativo, se torna positivo ao ser analisado em conjunto com seu par. Por outro lado, se é essa a conclusão a que se chega, ambos os indicadores deveriam ser somados para se ter uma medida

de quantos novos artigos foram publicados.

Ainda sobre impactos negativos, na concessão de bolsas de doutorado pelo CNPq encontrou-se impacto negativo nos seguintes indicadores: Orientação em Andamento de Graduação, Participação em Banca de Graduação e Outras Orientações Concluídas. O resultado é provável e faz sentido: na medida em que o professor se doutora, passa a ter mais contato com PPGs e menos contato com alunos de graduação, se especializando e orientando pesquisas mais complexas. O motivo apresentado está de acordo com o impacto positivo em novos projetos de pesquisa. No entanto, não há impacto positivo em orientações de mestres e doutores. Assim, pode-se supor, por exemplo, que a janela de tempo ainda não foi o suficiente para que o recém-doutor se firme e comece a orientar nesses níveis mais especializados. Conclui-se que há impacto negativo em “orientações”, mesmo que temporário, que pode valer a pena ser estudado.

A mesma observação deve ser realizada para todos os impactos que possuem clara relação um com o outro, como, por exemplo, todos os que tem “Início” e “Fim” em seus nomes. Por exemplo, na Tabela 7, encontrou-se o impacto de “Novos Vínculos” em 6 para o caso do Universal, mas não foi encontrado qualquer impacto em “Fim de Vínculos”. Portanto, pode-se interpretar que houve de fato novos vínculos como impacto da chamada. Isso não é verdade para outros resultados encontrados, em que Início de Vínculo deve ser subtraído de Fim de Vínculo para se encontrar Novos Vínculos. Dessa forma, pretende-se atualizar o programa de tratamento de dados disponibilizado no GitHub para que também agrupe indicadores, permitindo melhor interpretação desses.

11.4. Limitações e Estudos Futuros

Uma das limitações deste estudo foi a ausência da dimensão da qualidade da pesquisa, uma vez que se espera que o financiamento público afete diretamente essa dimensão, assim como restou comprovado a correlação com a dimensão da produtividade. Dessa forma, para a continuidade deste estudo, pretende-se buscar forma automatizada de trazer o número de citações de cada artigo encontrado nos Currículos Lattes analisados. Algumas primeiras tentativas já foram realizadas nesse sentido, sem sucesso. Mas, não foi realizada ampla pesquisa, de modo que isso ainda será feito e a solução, se possível, levará a outro estudo.

Outra limitação foi a ausência de análise de redução de dimensões. Nota-se que os 140 indicadores adotados no Lattes se correlacionam entre si e, para efeitos de uso em modelos preditivos ou *inforgraphics* para subsídio na avaliação de pesquisadores, podem ser reduzidos

à apenas algumas dimensões, como “produtividade”, “qualidade”, “formação humana”, “produção técnica”, etc. Esse é mais um estudo que se seguirá à esta tese.

Para impactos como editais de INCTs e de Pesquisadores Produtividade, pode-se criar ferramenta que busca coautorias, colaborações e correlações, de modo a colocar no grupo de tratamento todos aqueles beneficiados, direta ou indiretamente, com o financiamento. Dessa forma, o impacto poderia ser medido, não só para o pesquisador Coordenador ou PQ, mas para todos os que são por eles liderados. A ferramenta permitiria estender a análise do impacto a praticamente qualquer política pública de investimento em ciência cujos participantes usam o Currículo Lattes como principal ferramenta de registro profissional.

Foi realizada aqui análises de impacto e de custo-efetividade, mas não foi realizada análise de eficiência, pois não foi possível valorizar os diversos indicadores presentes no Lattes. Ou seja, para uma análise de eficiência, seria necessário estudo indicando qual o valor de um artigo publicado ou de um novo vínculo formado. Esse estudo nos parece ser teoricamente impossível de ser realizado devido às limitações da própria natureza da ciência já indicadas neste trabalho (não individualizável, temporalidade imprevisível e alta, de difícil monetização). Dessa forma, não há a pretensão de buscar avanços na custo-efetividade da ciência realizada.

11.5. Considerações Finais

Esta tese apresentou um método para calcular o Impacto de uma Política Pública de Financiamento em Ciência, de forma a ser escalável a várias políticas públicas, transparente, que permita comparação entre avaliações diferentes, tanto de forma longitudinal (entre políticas) quanto vertical (no tempo) e adequada a usos frequentes e automatizados ou semiautomatizados. O modelo busca inferir e medir a causalidade entre um dos diversos indicadores retirados do Currículo Lattes, comparando esse indicador entre duas distintas datas e levando-se em consideração o investimento realizado. Para isso, foi desenvolvido modelo quase-experimental de análise econométrica baseada no método das Diferenças em Diferenças e com o uso de múltiplas regressões lineares multivariadas.

Para colocar o método em prática, foram baixados e tratados sete milhões de Currículos Lattes, num contexto de *Big Data*, com o uso da linguagem Python e do banco de dados PostgreSQL. Utilizando-se do currículo baixado e descompactado, foi realizada varredura no arquivo XML por meio das definições presentes no arquivo XSD presente em <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>. Em termos de *Data Science*, cada

atributo em que “ANO” estivesse disponível, foi retirado um indicador, como por exemplo “Artigo Publicado” e “Orientação de Doutorado Concluída”. Dessa forma foram formados, neste primeiro estudo, 140 distintos indicadores, cuja lista completa pode ser pega, assim como todo o código fonte usado, em https://github.com/AlbertoCamposSilva/CV_Lattes_Analyzer.

Para testar o método proposto, ele foi aplicado em duas distintas políticas públicas de investimento à ciência e os resultados estudados. No entanto, no processo de se estudar, percebeu-se possibilidade de presença de misoginia no CNPq antes nunca detectada ou estudada. Dessa forma, essa misoginia também foi estudada, encontrando-se forte indicador de presença de discriminação sexual contra as mulheres, presente principalmente no valor financiado a menor para mulheres e no teto de vidro para cargos de alto *status* ou poder.

Encontrou-se também que a CAPES e o CNPq possuem impactos quase equivalentes em suas políticas de concessão de bolsas de doutorado, mas que o financiamento no Universal possui um impacto ainda maior. Comprovou-se, portanto, que a ferramenta pode ser aplicada e que traz medida do impacto do investimento público em ciência, podendo ser amplamente adotada e até mesmo automatizada.

12. REFERÊNCIAS BIBLIOGRÁFICAS

ALBERTO BOSCHETTI; LUCA MASSARON. **Python Data Science Essentials**. 2nd Edition ed. [s.l.: s.n.].

ALVES, M. R.; BARBOSA, M. C.; LINDNER, E. L. Diversidade e percepção de igualdade de gênero nos cursos de ciências exatas da UFRGS. **Revista Brasileira de Ensino de Ciência e Tecnologia**, v. 16, n. 1, 2023.

ANA PAULA ALVES PEREIRA et al. **Exatas é com elas: tecendo redes**. FUNDAÇÃO CECIERJ ed. [s.l.: s.n.].

ANGRIST, J. D.; PISCHKE, J.-S. **Mostly harmless econometrics: An empiricist's companion**. [s.l.] Princeton university press, 2009.

Apresentação — Conselho Nacional de Desenvolvimento Científico e Tecnológico. Disponível em: <<https://www.gov.br/cnpq/pt-br/acesso-a-informacao/institucional/institucional>>. Acesso em: 5 abr. 2023.

AREAS, R. et al. Gender and the scissors graph of Brazilian science: from equality to invisibility. 2020.

ARÊAS, R.; SANTANA, A. E.; BARBOSA, M. C. A discriminação da intelligentsia no campo científico brasileiro. **Revista Educação, Cultura e Sociedade**, v. 10, n. 3, 2020.

BEZERRA, A. R. G.; STANISCUASKI, F.; BARBOSA, M. C. Brazilian research productivity fellows in physics and nursing under a gender lens: 17 years of data. **Cadernos de Gênero e Tecnologia**, v. 46, n. 15, p. 64–79, 2022.

BIOLOGY, A. S. FOR C. **DORA: San Francisco Declaration on Research Assessment**. , 2012.

BORGES DE LIMA FACÃO, J. et al. **Energia & Gênero: Avaliação sobre Igualdade de Gênero no Setor de Energia do Brasil**. 10º CONGRESSO BRASILEIRO DE PESQUISA E DESENVOLVIMENTO EM PETRÓLEO E GÁS. **Anais...2019**. Disponível em: <<http://dx.doi.org/10.13140/rg.2.2.19253.52964>>. Acesso em: 8 maio. 2023

BORNMANN, L. Scientific peer review. **Annual Review of Information Science and Technology**, v. 45, p. 197–245, 2011.

BORNMANN, L. What is societal impact of research and how can it be assessed? A literature survey. **Journal of the American Society for Information Science and Technology**, v. 64, n. 2, p. 217–233, 2013a.

BORNMANN, L. What is societal impact of research and how can it be assessed? a literature survey. **Journal of the American Society for Information Science and Technology**, v. 64, n. 2, p. 217–233, fev. 2013b.

BORNMANN, L. How to analyze percentile citation impact data meaningfully in bibliometrics: The statistical analysis of distributions, percentile rank classes, and top-cited papers. **Journal of the American Society for Information Science and Technology**, v. 64, n. 3, p. 587–595, mar. 2013c.

BRASIL. **Constituição.** Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 22 set. 2022.

BRITO, C. et al. Harassment in Brazilian universities: How big is this problem? The Federal University of Rio Grande do Sul (UFRGS) as a case study. **Anais da Academia Brasileira de Ciências**, v. 94, n. 2, 2022.

CABRERA, M. R.; ARAÚJO, M. S. T. DE. Análise das políticas educacionais na Pós-graduação Stricto-Sensu no âmbito da formação continuada de professores. **Revista de Ensino de Ciências e Matemática**, v. 9, n. 3, p. 21–37, 28 jun. 2018.

CAPES. **Competências.** Disponível em: <<http://www.capes.gov.br/acessoainformacao/80-conteudo-estatico/acesso-a-informacao/5418-competencias>>. Acesso em: 8 ago. 2016.

CNPQ. **Edital MCT/CNPq N° 014/2010 - Universal.** Disponível em: <http://cnpq.br/chamadas-publicas?p_p_id=resultadosportlet_WAR_resultadoscnpqportlet_INSTANCE_0ZaM&filtro=encerradas&detalha=chamadaDivulgada&idDivulgacao=388>. Acesso em: 16 ago. 2016.

CNPQ. **CHAMADA INCT – MCTI/CNPq/CAPES/FAPs nº 16/2014.** Disponível em: <<http://www.cnpq.br/documents/10157/b91b7566-2110-4a29-9704-88cdd324e072>>. Acesso em: 16 ago. 2016.

CNPQ. **CNPq - BOLSAS INDIVIDUAIS NO PAÍS.** Disponível em: <http://www.cnpq.br/web/guest/view/-/journal_content/56_INSTANCE_0oED/10157/2958271>. Acesso em: 16 ago. 2016.

CNPQ. **Apresentação — Português (Brasil).** Disponível em: <<https://www.gov.br/cnpq/pt-br/acesso-a-informacao/institucional/institucional>>. Acesso em: 23 set. 2022.

CNPQ. **CNPq - Acesso à Informação - Dados Abertos - Consulta de Dados.** Disponível em: <http://dadosabertos.cnpq.br/pt_BR/>. Acesso em: 23 set. 2022.

COE, I. R.; WILEY, R.; BEKKER, L.-G. Organisational best practices towards gender equality in science and medicine. **The Lancet**, v. 393, n. 10171, p. 587–593, 9 fev. 2019.

COLWELL, R. et al. Informing research choices: Indicators and judgment. **The Expert Panel on Science Performance and Research Funding**, p. 142, 2012.

Competências — Português (Brasil). Disponível em: <<https://www.gov.br/cnpq/pt-br/acesso-a-informacao/institucional/competencias>>. Acesso em: 22 jun. 2022.

CORAK, M. **Achieving Inclusive Growth in the Asia Pacific.** [s.l.] ANU Press, 2020.

CÓSER, I. et al. Determinants of patents in the life sciences and health sciences at federal universities in Minas Gerais state, Brazil: An analysis of panel data for 1995-2016. **Cadernos de Saude Publica**, v. 34, n. 6, 2018.

DA SILVA, R. P. A. et al. Misogyny in brazilian federal government agencies for science and high-education. **Anais da Academia Brasileira de Ciências**, v. 93, 2021.

DE OLIVEIRA SANTIAGO, M.; AFFONSO, F.; MAGELA RODRIGUES DIAS, T. Perfil das orientações e produções das mulheres fundamentado em dados da Plataforma Lattes. **Ciência da Informação**, v. 49, n. 3, p. 188–203, 2020.

DER MEULEN, B.; RIP, A. Evaluation of societal quality of public sector research in the Netherlands. **Research evaluation**, v. 9, n. 1, p. 11–25, 2000.

Doctorate Recipients from U.S. Universities: 2021 | NSF - National Science Foundation. Disponível em: <<https://nces.nsf.gov/pubs/nsf23300/data-tables>>. Acesso em: 7 abr. 2023.

DONOVAN, C. The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research. **New directions for evaluation**, v. 2008, n. 118, p. 47–60, 2008.

DOUGLAS, K. **Energy subsidies and the environment. Subsidies and environment: exploring the linkages.** OECD, , 1996.

E3M. Green Paper: Fostering and Measuring “Third Mission” in Higher Education Institutions. p. 23, 2012.

Extração de dados - Portal Memória. Disponível em: <<http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>>. Acesso em: 5 jul. 2022.

FERRARI, N. C. et al. Geographic and gender diversity in the Brazilian Academy of Sciences. **Anais da Academia Brasileira de Ciências**, v. 90, n. 2, p. 2543–2552, 2018.

FORSCHUNGSGEMEINSCHAFT, D. Proposal Preparation Instructions. **DFG form**, v. 54, 2012.

GARCÍA-PEÑALVO, F. J. et al. **Women in STEM in Higher Education.** Singapore: Springer Nature Singapore, 2022.

GERTLER, P. J. et al. **Avaliação de impacto na prática.** Segunda Ed ed. [s.l.] World Bank Publications, 2018.

GIMENES, S. C.; CARVALHO, M. J. S. Avaliação de uma política pública de inclusão digital : os telecentros comunitários de porto alegre. 2008.

GOULDER, L. H.; PARRY, I. W. H. H. Instrument choice in environmental policy. **Review of environmental economics and policy**, v. 2, n. 2, p. 152–174, 2008.

GREENLAW, S. A. et al. **Principles of economics 2e.** [s.l.: s.n.].

GRUS, J. Data science from scratch: first principles with python. 2019.

GUTHRIE, S. et al. Measuring Research: A Guide to Research Evaluation Frameworks and Tools. **RAND Monographs**, p. 189, 2013.

História e missão — CAPES. Disponível em: <<https://www.gov.br/capes/pt-br/aceso-a-informacao/institucional/historia-e-missao>>. Acesso em: 5 abr. 2023.

International Day of Women and Girls in Science | United Nations. Disponível em: <<https://www.un.org/en/observances/women-and-girls-in-science-day>>. Acesso em: 10 out. 2022.

KINCAL, G. et al. Cross Border Business Cycle Impacts on the El Paso Housing Market. **Estudios fronterizos**, v. 11, n. 22, p. 49–77, 2010.

LEYDESDORFF, L.; IVANOVA, I. A. “Open Innovation” and “Triple Helix” Models of Innovation: Can Synergy in Innovation Systems Be Measured? **Available at SSRN 2791914**, 2016.

LI, J.; DOSTIE, B.; SIMARD-DUPLAIN, G. Firm Pay Policies and the Gender Earnings Gap: The Mediating Role of Marital and Family Status. **ILR Review**, p. 001979392210935, 30 abr. 2022.

LINDSEY, D. Using citation counts as a measure of quality in science measuring what’s measurable rather than what’s valid. **Scientometrics**, v. 15, p. 189–203, 1989.

LUGOBONI, L. F. O Currículo Lattes. **Revista Liceu On-Line**, v. 10, n. 2, p. 1–5, 14 jul. 2020.

MCTI, M. DA C. T. E I. Plano Estratégico do Ministério da Ciência, Tecnologia e Inovações - PEI MCTI 2020-2023. ago. 2022.

MELOROSE, J.; PERROY, R.; CAREAS, S. The Australian Research Quality Framework: A Live Experiment in Capturing the Social, Economic, Environmental and Cultural Returns of Publicly Funded Research. **Statewide Agricultural Land Use Baseline 2015**, v. 1, p. 1–21, 2015.

MILESI, C. et al. Charting the Impact of Federal Spending for Education Research: A Bibliometric Approach. **Educational researcher**, v. 43, n. 7, p. 361–370, 2014.

MOED, H. F.; HALEVI, G. Multidimensional assessment of scholarly research impact. **Journal of the Association for Information Science and Technology**, v. 66, n. 10, p. 1988–2002, 1 out. 2015.

MÖLLER, T.; SCHMIDT, M.; HORNBOSTEL, S. Assessing the effects of the German Excellence Initiative with bibliometric methods. **Scientometrics**, v. 109, n. 3, p. 2217–2239, 2016.

MORAIS, E. L. S. DE et al. O Mestrado Profissional em Ensino de Ciências Naturais e Matemática da UFRN: uma análise das dissertações e acompanhamento de egressos de 2012 a 2020. **Revista de Ensino de Ciências e Matemática**, v. 13, n. 3, p. 1–19, 12 maio 2022.

MÜNCH, R.; SCHÄFER, L. O. Rankings, Diversity and the Power of Renewal in Science. A Comparison between Germany, the UK and the US. **European Journal of Education**, v. 49, n. 1, p. 60–76, 2014.

NODARI, C. A.; SOARES DE OLIVEIRA, C. C.; BARBOSA, M. C. **A Comparison of the Professional Profile of Men and Women Deans of Brazilian Universities** *Revista Brasileira de Pós-graduação (RBPG)*. [s.l.: s.n.]. Disponível em: <<http://www.if.ufrgs.br/~barbosa/gender.html>>. Acesso em: 8 maio. 2023.

OCDE. **Manual de Canberra**. Paris: OCDE, 1996.

OCDE. **Manual de Frascati 2002: Medição de atividades científicas e tecnológicas**. [s.l.: s.n.].

OLIVEIRA, A. et al. Gênero e desigualdade na academia brasileira: uma análise a partir dos bolsistas de produtividade em pesquisa do CNPq. <http://journals.openedition.org/configuracoes>, n. 27, p. 75–93, 17 jun. 2021.

OLIVEIRA, A. et al. O perfil dos bolsistas de produtividade em pesquisa do CNPq em Sociologia. **Sociologias**, v. 24, n. 59, p. 170–198, jan. 2022.

OLIVEIRA, K. H.; SERRA, M. M. Mulheres, tempos e espaços na ciência agropecuária paulista. **Revista Ártemis**, v. 25, n. 1, p. 203–203, 2018.

OLIVEIRA, T.; AMARAL, L. Public Policies in Science and Technology in Brazil: challenges and proposals for the use of indicators in evaluation. **Mugnaini, R**, p. 189–217, 2017.

Oslo Manual 2018. [s.l.] OECD, 2018.

PERLIN, M. S.; SANTOS, A. A. P. Os pesquisadores, as publicações e os periódicos da área de Finanças no Brasil: Uma análise com base em currículos da plataforma Lattes. **Revista brasileira de finanças**, v. 13, n. 2, p. 162–199, 2015.

PERUCCHI, V.; MUELLER, S. P. M. Estudo com as patentes produzidas e o perfil dos inventores dos Institutos Federais de educação, ciência e tecnologia. **Rev. digit. bibliotecon. cienc. inf.**, v. 12, n. 1, p. 191–213, jan. 2014.

PINDYCK, R. S.; RUBINFELD, D. L. Microeconomia. 7ª edição. São Paulo. Ed. **Saraiva, São Paulo**, 2010.

PINO, F. A. A Questão da Não Normalidade: uma Revisão. **Revista de economia agrícola**, v. 61, n. 2, p. 17–33, 2014.

PNE - Plano Nacional de Educação - Plano Nacional de Educação - PNE. Disponível em: <<https://pne.mec.gov.br/>>. Acesso em: 5 abr. 2023.

Portal CNPq - RN-028/2015. Disponível em: <http://memoria2.cnpq.br/view/-/journal_content/56_INSTANCE_0oED/10157/2958271>. Acesso em: 22 out. 2022.

PRESIDÊNCIA DA REPÚBLICA; MINISTÉRIO DA EDUCAÇÃO; SECRETARIA DE ARTICULAÇÃO COM OS SISTEMAS DE ENSINO. **Planejando a Próxima Década**. [s.l.: s.n.].

Proposal and Award Policies and Procedures Guide, January 2016. 25 jan. 2016.

PSACHAROPOULOS, G.; PATRINOS, H. A. Returns to investment in education: a decennial review of the global literature. **Education Economics**, v. 26, n. 5, p. 445–458, 3 set. 2018.

RESENDE, C. C. DE et al. O impacto da educação integral na participação das mães no mercado de trabalho e no trabalho infantil: Uma avaliação de impacto do Programa Mais Educação por regressão descontínua. **Revista Brasileira de Ciência Política**, n. 32, p. 323–362, ago. 2020.

REZENDE, F. DA C. Transformações na cientificidade e o ajuste inferencial na Ciência Política: Argumento e evidências na produção de alto fator de impacto. **Revista de Sociologia e Política**, v. 25, n. 63, p. 103–138, 2017.

RIGONATTO, C. A. Quem paga a conta?: subsídios e reserva legal: avaliando o custo de oportunidade do uso do solo. p. 120, 2006.

ROCHA-CARPIUC, C.; MADEIRA, R. M. Gender inequality, internationalization and academic trajectories in Political Science Evidences from Brazil and Uruguay. **Civitas**, v. 19, n. 3, p. 545–563, 15 dez. 2020.

RODRIGUES, A. A.; DIAS, G. A. Estudos sobre visualização de dados científicos no contexto da Data Science e do Big Data. **Pesquisa Brasileira em Ciência da Informação. João Pessoa.**, v. 12, n. 1, p. 219–228, 2017.

RODRIGUES, J. G.; GUIMARÃES, M. C. S. A Fundação Oswaldo Cruz e a ciência no feminino: A participação feminina na prática e na gestão da pesquisa em uma instituição de ensino e pesquisa. **Cadernos Pagu**, v. 2016, n. 46, p. 197–222, 1 jan. 2016.

ROSA ALVES, M.; CRISTINA BERNARDES BARBOSA, M.; LUIZ LINDNER, E. **Diversity and perception of gender equality in exact science programs at UFRGS ABSTRACT** *Brazilian Journal of Science Teaching and Technology*. [s.l: s.n.]. Disponível em: <<https://periodicos.utfpr.edu.br/rbect>>.

ROSA ALVES, M.; LINDNER, E. L.; BARBOSA, M. C. **A percepção de gênero, trabalho e ciência entre estudantes dos anos iniciais do ensino fundamental**. [s.l: s.n.]. Disponível em: <<http://www.if.ufrgs.br/~barbosa/Publications/Gender/alves-apercepcao-2023.pdf>>. Acesso em: 8 maio. 2023.

SACCARO, A. **Ampliação do ensino superior brasileiro: um estudo sobre as causas da evasão e o impacto da Bolsa Permanência do PNAES**. Dissertação—[s.l.] Pontifícia Universidade Católica do Rio Grande do Sul, 9 fev. 2016.

SALDANHA, R. DE F. A pesquisa científica na era do Big data: cinco maneiras que mostram como o Big data prejudica a ciência, e como podemos salvá-la. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde**, v. 16, n. 3, p. 742–745, 30 set. 2022.

SALTER, A. J.; MARTIN, B. R. The economic benefits of publicly funded basic research: a critical review. **Research policy**, v. 30, n. 3, p. 509–532, 2001.

SILVA, A. DE C. E. **Análise Custo-Efetividade dos Investimentos em Meio Ambiente no Edital MCT/CNPq nº 14/2010 - UNIVERSAL**. Dissertação—Brasília: Universidade de Brasília, 4 jun. 2018.

SILVA, V. C. DA; PORCIÚNCULA, L. DE O. Impactos da pós-graduação stricto sensu na formação de professores de português da educação básica do Distrito Federal. **Formação Docente – Revista Brasileira de Pesquisa sobre Formação de Professores**, v. 9, n. 17, p. 141–162, 31 dez. 2017.

SILVA, W. DE M. F. Consórcios públicos na gestão de resíduos sólidos urbanos no Brasil. 2015.

- SIMONIS, U. E. et al. Environmental policy instruments in a new era. v. 01,404, n. Cml, p. 53, 2001.
- SOEIRO, A. et al. Green Paper. Fostering and Measuring Third Mission in Higher Education Institutions. 2012.
- SOEIRO, A.; MONTESINOS, P.; OTHERS. E3M-EUROPEAN INDICATORS AND RANKING METHODOLOGY FOR UNIVERSITY THIRD MISSION. 2011.
- SÔNIA DOMINGUES SANTOS BRAMBILLA; IDA REGINA CHITTÓ STUMPF. Produção científica da UFRGS representada na Web of Science (2000-2009). **Perspectivas em Ciência da Informação**, v. 17, p. 34–50, 2012.
- SPINAK, E. Indicadores cientimetricos. **Ciência da informação**, 1998.
- TEIXEIRA, L. I. L. et al. Mestrados profissionais e perspectivas acadêmicas para servidores técnico-administrativos. **Revista Gestão Universitária na América Latina - GUAL**, p. 276–291, 20 dez. 2018.
- TODESCHINI, R.; BACCINI, A. **Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research**. [s.l.] John Wiley & Sons, 2016.
- VSNU, K.; OTHERS. Standard Evaluation Protocol 2009-2015. 2009.
- VSNU; NWO; KNAW. Standard Evaluation Protocol 2015-2021. 2015.
- WILBERTZ, J. Evaluating societal relevance of research. 2013a.
- WILBERTZ, J. Evaluating the societal relevance of research. n. april, 2013b.

13. ANEXO I - ARTIGO: INFERINDO A CAUSALIDADE ENTRE FINANCIAMENTO EM CIÊNCIAS E A FORMAÇÃO DE RECURSOS HUMANOS: UM MODELO ECONOMETRICO PARA MEDIR O IMPACTO DE POLÍTICAS PÚBLICAS POR UM QUASE-EXPERIMENTO

INFERRING CAUSALITY BETWEEN SCIENCE FUNDING AND HUMAN RESOURCE TRAINING: AN ECONOMETRIC MODEL TO MEASURE THE IMPACT OF PUBLIC POLICIES THROUGH A QUASI-EXPERIMENT

[Alberto de Campos e Silva](#)

Universidade Federal do Rio Grande do Sul
silva@albertocampos.com.br

[Alexandre Guilherme Motta Sarmento](#)

Universidade Federal do Rio Grande do Sul
agmottas@gmail.com

RESUMO

Este estudo buscou método para inferir a correlação entre uma política pública de investimento em ciência e vários indicadores presentes no Currículo Lattes do CNPq. A inferência é realizada por meio de um quase-experimento, em um modelo econométrico baseado no método das diferenças em diferenças e variáveis instrumentais, quando possível, ou pareamento nos demais casos. Dessa forma, obtém-se uma medida do Impacto nas Políticas Públicas de Financiamento em Ciência com uso de indicadores presentes no Lattes, incluindo publicação de livros e revistas, desenvolvimento de materiais escolares, formação humana nos níveis de mestrado e doutorado, artigos publicados, patentes, entre diversos outros. O método proposto é exploratório, de abordagem quantitativa, escalável, transparente, que permite comparação entre avaliações diferentes e adequado a uso frequente e longitudinal. Por outro lado, não é descritivo, flexível nem permite inferir causas ou demais efeitos e, portanto, inadequado como ferramenta exclusiva de escolha. Assim, são apresentadas as ferramentas para a análise de impacto, como escolher um modelo econométrico, como aplicar a cientometria no Lattes, como aplicar o método das diferenças em diferenças, chegando-se ao modelo proposto e suas limitações.

PALAVRAS-CHAVE: Avaliação de Impacto, Cientometria, Modelo Econométrico, Currículo Lattes, Indicadores Bibliométricos.

13.1. Introdução

Este artigo apresenta instrumentos bem como busca propor um método para a avaliação

de impacto de uma política pública de investimento em ciências. Segundo Bornmann (2013b), o impacto da ciência é a avaliação do retorno social, cultural, ambiental e econômico. No entanto, observa-se que as políticas públicas brasileiras, principalmente aquelas que implicam em transferência de recursos, possuem falta de cultura de avaliação de impacto, desde seu desenho até sua evolução (BORNMANN, 2013a).

Muitos programas não realizam avaliações piloto ou mesmo avaliações rigorosas do programa alguns anos após sua implementação. Além disso, as avaliações, quando realizadas, são comumente focadas apenas nos insumos (como, por exemplo, o valor financiado). Ocorre que “A utilização de métricas de insumo para “avaliar” o programa cria um incentivo natural tanto para iniciá-lo em larga escala como para promover o seu crescimento acelerado”, mas sem a correta avaliação de sua efetividade, eficiência ou impacto (RESENDE et al., 2020). Isso porque a avaliação do insumo costuma apontar uma alta eficácia (100% dos recursos investidos), mesmo se houver baixo impacto.

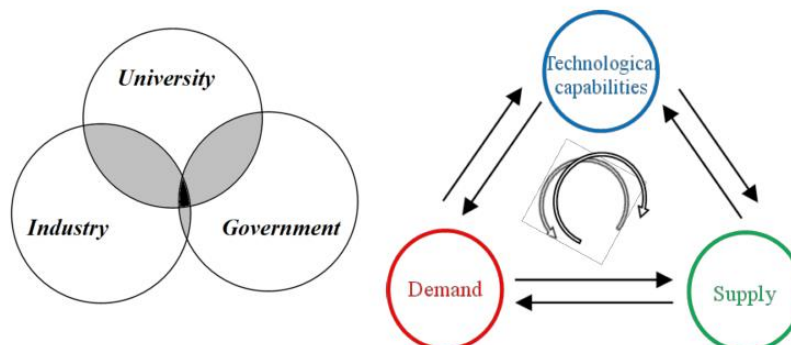
A expressão “Modo 2” de análise da ciência foi criada por Gibbons et al. (1994). Mas, Bornmann (2013b) expande o conceito e denomina de Modo 2 as novas formas de análise da ciência, realizada “levando-se em conta considerações alternativas de mudanças em curso na prática científica, como o ‘capitalismo acadêmico’, ‘ciência pós-normal’, ‘Triple Helix’, ‘universidade empresarial’, ‘a ciência pós-acadêmica’, ‘pesquisa estratégica’ e ‘ciência de finalização’”.

As novas formas de avaliar a ciência devem levar em consideração as interações entre a academia, empresas e sociedade, além da excelência e originalidade, levando-se em conta valores e critérios de utilidade prática, avaliando-se o impacto social, além da qualidade científica (BORNMANN, 2013a). Assim, a avaliação deve considerar a relação entre a academia, a esfera governamental, social e econômica (indústria e comércio), como uma forma de relações em três dimensões (Figura 23). A adição da academia na Tripla Hélice torna o progresso dependente do caminho, de modo que cada eixo tanto impeça quanto reforce alterações nos outros dois vértices, em um sistema dinâmico e autogerido (Figura 24) (LEYDESDORFF; IVANOVA, 2016). Assim, não basta que o governo invista, de modo independente, na ciência, na indústria ou no comércio, mas o investimento deve ser realizado de modo a fazer com que os três se reforcem e promovam mutuamente o progresso do país. Da mesma forma, a análise do impacto deve buscar medir essa interação.

Nesse sentido, Wilbertz (2013) define o impacto da pesquisa pública como “a medida do retorno social, cultural, ambiental e econômico de uma pesquisa financiada pelo setor público, sejam produtos ou ideias”. Já Bornmann (2013) define o impacto da pesquisa pública

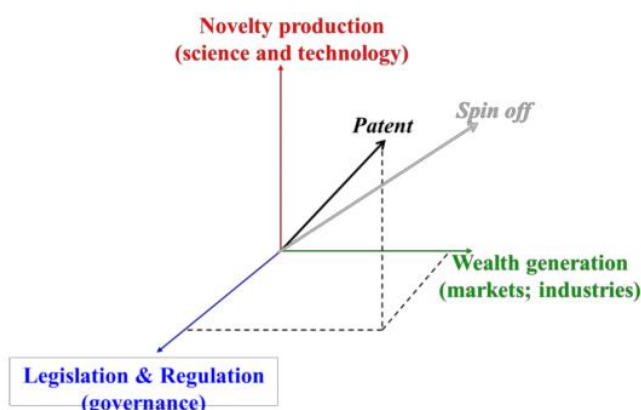
como “qualquer análise sistemática e baseada em dados com o objetivo de determinar ou prever os impactos sociais e econômicos da pesquisa ou das atividades técnicas correlatas” (tradução nossa).

Figura 23 – Representações da Tripla Hélice em termos de um Diagrama de Venn ou como triângulo.



Fonte: ETZKOWITZ & LEYDESDORFF, 2000, p. 111, Fig. 1a e PETERSEN et al., 2016, p. 667, Fig. 1b, ambos *apud* LEYDESDORFF; IVANOVA, 2016.

Figura 24 – A Hélice Tripla como um sistema cartesiano de Coordenadas



Fonte: IVANOVA, LEYDESDORFF, 2014. *apud* LEYDESDORFF; IVANOVA, 2016.

Esta pesquisa propõe um método de avaliação entre o eixo de governança da tripla hélice com o eixo científico, apresentando um método de inferir o impacto de uma política pública de investimento em ciência não só na dimensão da qualidade da ciência, mas também no eixo cultural e social. Além desta introdução, este artigo se subdivide em Revisão Bibliográfica, Modelo Econométrico e Conclusão. A Revisão Bibliográfica é apresentada conforme foram escolhidas as Ferramentas para a Análise do Impacto e conforme é proposto como aplicar estas ferramentas de análise de impacto em estudo cientométrico com o método quase-experimental e análises econométricas. O modelo econométrico e a forma de cálculo do impacto da pesquisa são então propostos no Capítulo 3, com a conclusão ao final apresentada no Capítulo 4.

13.2. Revisão Bibliográfica

13.2.1. Ferramentas para a Análise do Impacto

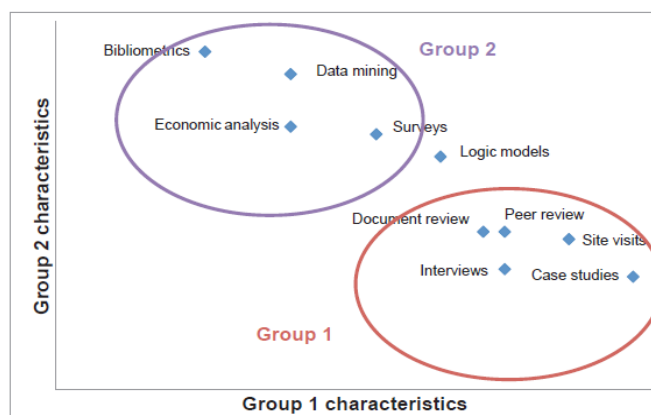
A avaliação possui diversos significados, a depender do contexto. De qualquer forma, a avaliação é sempre indicada como um:

(...) procedimento estratégico e imprescindível para dar transparência às ações públicas, democratizar o Estado e a sociedade civil, a fim de compreender o Estado em ação através das políticas públicas, melhorando-as e sugerindo modificações durante a sua formulação, a sua implementação e também nos resultados. (...) Sendo assim podemos dizer que a avaliação é um instrumento para maximizar a eficácia e eficiência dos programas sociais

(GIMENES; CARVALHO, 2008).

Guthrie et al. (2013) indicam que as ferramentas utilizadas nas análises podem ser agrupadas como “Formativas” (Grupo 1, mais flexíveis, normalmente usam de julgamento e interpretação) e como “Quantitativas” (Grupo 2, *escaláveis, transparentes, permitem comparações entre avaliações diferentes e são adequadas a usos frequentes e longitudinais*).

Figura 25 – Ferramentas de Análise da Pesquisa pertencentes ao Grupo 1 e 2



Fonte: GUTHRIE, S. et al. *Measuring Research: A Guide to Research Evaluation Frameworks and Tools*. RAND Monographs, p. 10, 2013.

Segundo Bornmann (2013), a melhor forma de avaliar o impacto do investimento em P&D é por meio de análises sistemáticas, quantitativas e contínuas. No entanto, faltam avaliações de impacto de políticas públicas, principalmente às relacionadas ao financiamento a P&D, isso porque o impacto é de difícil mensuração, mesmo com o uso das ferramentas do Grupo 2 acima indicadas. De qualquer forma, para o cálculo do impacto busca-se uma avaliação contínua, sistemática e quantitativa e, para tal, as ferramentas do Grupo 2 foram consideradas mais apropriadas: Bibliometria (ou cientometria), Análise Econométrica e *Data Mining*. Passemos à análise de como aplicar essas ferramentas neste estudo.

13.2.2. Indicadores para a Análise Econométrica

O modelo econométrico busca analisar a correlação entre uma variável dependente e suas variáveis independentes. Assim, para aferir o “impacto” por meio de um modelo econométrico de regressão linear, busca-se estimar os parâmetros de cada variável independente de modo a minimizar os erros e, com isso, estimar o impacto total. Dessa forma, as variáveis independentes são nada mais que indicadores, que buscam representar uma realidade.

No entanto, não há consenso sobre qual indicador usar. O Manual de Frascati 2002 (OCDE, 2002) busca apresentar métodos para a medição de atividades científicas e tecnológicas. O Manual reconhece a dificuldade em se definir um sistema de indicadores para avaliar CT&I, reconhecendo que não existe um modelo único capaz de estabelecer relações causais entre a ciência, a tecnologia, a economia e a sociedade. O Manual de Oslo (Oslo Manual 2018, 2018) oferece métodos para a compilação de dados que buscam inferir a inovação dentro da Ciência e Tecnologia. Em seu anexo I, apresenta métodos e casos relatados em periódicos técnicos e comerciais, o que surpreende pela falta de menção a artigos científicos sobre a questão. O Manual de Canberra (OCDE, 1996) buscou oferecer orientações para a medida de recursos humanos em CT&I, bem como sobre a análise desses dados. No entanto o manual nem sequer menciona métodos econométricos baseados em bibliometria.

Os exemplos acima indicam a falta, na literatura, de um método econométrico amplamente aceito. Este estudo sugere modelo econométrico capaz de verificar inferência causal entre um indicador em um determinado ano e indicadores de anos anteriores. No entanto, todas as medidas quantitativas de pesquisa são, na melhor das hipóteses, apenas indicadores parciais que sempre são influenciados em parte por outros fatores além do progresso científico que busca medir. Mesmo assim, de acordo com Moed et al. (2015), o uso seletivo e cuidadoso de tais indicadores é certamente melhor do que nenhum. Além disso, é provável que a abordagem mais frutífera envolva o uso combinado de vários indicadores.

São poucos os estudos econométricos que envolvam a bibliometria e, em sua maioria, envolvem apenas uma dimensão ou aspecto, como a quantidade de publicações ou a relação de autoria. No entanto, o impacto deve ser calculado em todas as suas dimensões econômicas, sociais, culturais e ambientais (BORNEMANN, 2013a). Por isso, será necessário usar indicadores para representar todas essas dimensões.

A hipótese levantada é de que a Plataforma de currículo Lattes possui dados que podem ser transformados em indicadores capazes de representar, em certa medida, todas essas dimensões. O modelo adotado deve ser aplicado a cada um desses indicadores, estimado o valor

do coeficiente que dá a medida da correlação entre o financiamento realizado e a variável dependente, e medindo se essa estimativa é real. Assim, temos a medida do impacto para cada um dos indicadores e, portanto, uma estimativa do impacto total da política pública.

Ocorre que estudos econométricos são dependentes do modelo econométrico adotado (ALBERTO BOSCHETTI; LUCA MASSARON, 2016). Esse é, provavelmente, um dos motivos pelos quais os pesquisadores evitam concluir por um impacto numericamente estabelecido, uma vez que a mudança no modelo pode implicar numa grande alteração no valor encontrado. A escolha do modelo adotado é, então, longe de ser algo simples ou direto e o modelo precisa ser explicado e logicamente aceito antes de eventual método estatístico aplicado, de modo que cada variável independente deve ser inclusa no modelo econométrico apenas se sua inclusão também fizer sentido teórico.

13.2.3. O Currículo Lattes e Data Mining

Data Mining é o processo de explorar grandes quantidades de dados à procura de padrões consistentes, com o uso de ferramentas computacionais (programas), extraindo ou ajudando a evidenciar padrões nestes dados e auxiliando na descoberta de conhecimento. Neste estudo, o Data Mining será aplicado no Currículo Lattes (Lattes).

O Currículo Lattes (Lattes) é uma plataforma virtual pertencente ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que apresenta numa mesma base de dados entradas para demonstrar a trajetória acadêmica e profissional, não só de pesquisadores, mas de qualquer profissional em atividade em qualquer área do conhecimento. O Lattes apresenta centenas de indicadores, suficientes para refletir aspectos sociais, culturais, ambientais e econômicos do pesquisador (LUGOBONI, 2020).

O preenchimento do Lattes é obrigatório em todos os editais públicos do CNPq e universalmente adotado como a principal ferramenta de currículo no Brasil. É estruturado (facilitando data mining), de acesso público e contempla mais de sete milhões de currículos cadastrados, além de ser usado por padrão em universidades, no governo (tanto federal, quanto estadual e municipal) e até mesmo por muitas empresas. Possui *webservice* próprio para a extração de currículos (Extrator Lattes) e definição de elementos estruturados em XML. Dessa forma, o Lattes se apresenta como ferramenta ideal para o estudo pretendido do Impacto da Pesquisa, por permitir Data Mining de forma excepcionalmente completa e precisa.

Portanto, é possível extrair do Lattes diversos indicadores, capazes de representar as diversas dimensões do impacto, considerando-se como indicador qualquer informação relevante que possa ser automaticamente extraída por um programa de computador e que

contenha a informação do ano de realização/acontecimento. Assim, podem ser considerados como indicadores, por exemplo: ano de conclusão do doutorado, da publicação de um artigo ou matéria de jornal, de uma premiação ou de uma obra de arte, de um vínculo ou de uma orientação.

Nota-se que foi realizado estudo prévio em que foram encontrados 140 distintos indicadores, cuja listagem completa pode ser encontrada em “Tabelas”, “indicadores_nomes”, disponível em https://github.com/AlbertoCamposSilva/CV_Lattes_Analyzer. No entanto, reforça-se que essa lista não é exaustiva e pode ser expandida, incluindo qualquer outra informação passível de ser extraída de forma automática.

13.2.4. Cientometria

A bibliometria é multidisciplinar e busca analisar a comunicação impressa, a forma mais importante de divulgação científica da academia (SPINAK, 1998). Já a cienciometria ou cientometria analisa aspectos econômicos e sociais das ciências físicas, naturais e sociais, examinando o desenvolvimento e as políticas da ciência. Aplicada de forma quantitativa, a cienciometria é considerada como disciplina ou atividade econômica e pode estabelecer comparações entre políticas públicas. Assim, a cienciometria usa técnicas matemáticas e análises estatísticas para investigar as características da pesquisa científica e pode ser considerada como um instrumento da sociologia da ciência (SPINAK, 1998).

Como esta pesquisa analisa um documento de texto, o Currículo Lattes, é um estudo bibliométrico. Mas, o estudo vai além, analisando informações contidas dentro do Currículo Lattes para inferir causalidade e medir impacto, de modo que este também é cientométrico. É importante mencionar que o Lattes apresenta listagem completa de cada artigo publicado por cada autor. Mas, aqui, a quantidade e a qualidade dos artigos publicados são apenas dois dos indicadores presentes, entre uma centena de outros.

13.2.5. Análise Econométrica

A análise econométrica busca medir a correlação entre uma variável independente com uma ou mais dependentes, de modo a inferir causalidade. Dessa forma, por meio de métodos estatísticos, sendo um dos mais usados a regressão linear ordinária pelos mínimos dos erros ao quadrado, mede-se a correlação entre essas variáveis. Pode-se mensurar a probabilidade de que essa correlação de fato exista, o quanto da variação da variável independente pode ser explicada pelas variáveis dependentes, entre outras, buscando inferir causalidade entre essas variáveis (GRUS, 2019).

Correlação é a dependência entre duas variáveis, que pode ser tanto positiva quanto negativa. Assim, por exemplo, podemos dizer que na infância há correlação entre peso e idade: quanto mais pesada for a criança, mais velha é. Observa-se, no entanto, que correlação não significa causalidade. No caso anterior, o peso depende da idade (e não o inverso), além de que o peso maior pode ser devido ao fato da criança engordar, entre inúmeros outros fatores. Assim, a inferência de causalidade por meio de medida estatística de correlação é algo que deve ser procedido com cautela e depende do modelo adotado.

São vários os exemplos de análises econométricas realizadas. Kincal et al. (2010) utilizaram de um modelo econométrico procurando evidência empírica sobre os impactos das flutuações do ciclo de negócios transfronteiriços nos mercados imobiliários metropolitanos localizados próximos às fronteiras internacionais examinando os impactos das condições econômicas no México sobre as vendas de casas unifamiliares existentes em El Paso, Texas. Ao encontrar que não há correlação estatisticamente significativa, Kincal et al. (2010) encontraram que evidências empíricas sólidas de tal ligação são elusivas.

Com relação ao uso de econometria e o Currículo Lattes, Cóser et al. (2018) estudaram a relação entre a publicação de patentes e indicadores no Lattes, como livros-texto, ementas de cursos, textos didáticos, artigos em periódicos científicos, livros e capítulos em livros, relatórios de pesquisa contratada, guias profissionais, artigos em jornais, comunicações submetidas a mídias sociais (incluindo blogs) e formação humana no grau de doutorado. Eles encontraram não haver correlação, ou até mesmo correlação negativa, entre esses indicadores e a publicação de patentes. O estudo de Cóser et al. (2018) surpreende, uma vez que se espera haver correlação entre uma pesquisa científica e a publicação de patentes.

13.2.6. Método das Diferença em Diferenças

Quando não é possível a seleção aleatória²⁷, o uso de variáveis instrumentais²⁸ ou o método da regressão descontínua²⁹, o método das diferenças em diferenças é o mais indicado.

²⁷ A seleção aleatória seria, *ceteris paribus*, escolher de forma totalmente aleatória quem seria financiado pela política pública. No caso de bolsas de doutorado, implicaria em se manter todo o processo de seleção, classificação e regras no edital, mas selecionar de forma totalmente aleatória os participantes ao final, algo que não é possível legalmente.

²⁸ O método se aplica quando há dependência entre os grupos de tratamento e de controle de alguma fonte externa de variação para determinar o status do tratamento. Intuitivamente, podemos descrever uma variável instrumental como algo fora do controle do indivíduo e que influencia sua probabilidade de participar de um programa, mas que não está associado a suas características. A seleção aleatória dentro de um grupo é um exemplo. No caso de políticas públicas de financiamento à ciência, o método é de difícil aplicação.

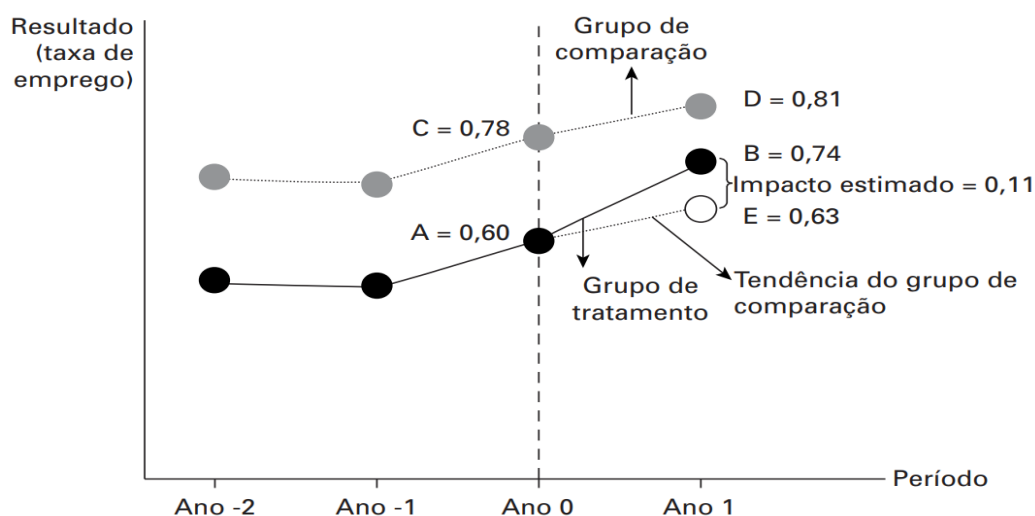
²⁹ Se houver um indicador com um ponto de corte bem definido, como uma nota ou classificação, pode-se pegar aqueles candidatos próximo a esse ponto de corte, beneficiados e não beneficiados, comparando-se os dois

O método é quase-experimental³⁰ e compara as mudanças nos resultados ao longo do tempo entre uma população que está inscrita em um programa (o grupo de tratamento) e uma população que não está (o grupo de controle).

No método, a diferença dos resultados no grupo de tratamento, subtraída da diferença dos resultados no grupo de controle, se traduziria então no impacto do tratamento, já desconsiderados os efeitos no tempo nesse grupo. O pressuposto é de que os efeitos no tempo dos dois grupos são equivalentes entre si. De forma mais específica, assume-se que a inclinação da reta no grupo de controle devido aos efeitos no tempo é igual a inclinação dessa mesma reta no grupo de tratamento (GERTLER et al., 2018).

Em outras palavras, o método pressupõe que não há diferenças entre os grupos que se alteram com o tempo, de modo que os grupos podem ser distintos entre si em características observáveis e não observáveis, mas essa distinção deve se manter constante no tempo para que o método seja válido. Portanto, a estimativa somente será enviesada se houver fatores que afetem a diferença das tendências entre os dois grupos que não foram considerados na regressão multivariada, mesmo que haja variáveis não observadas, como, por exemplo, financiamento externo ou dependência de outro fator.

Figura 26 – O método da Diferença em Diferenças



Observação: todas as diferenças entre pontos devem ser lidas como diferenças verticais nos resultados medidos no eixo vertical.

Fonte: Banco Mundial, Avaliação de Impacto na Prática, página 146 (GERTLER et al., 2018).

grupos, por meio de uma regressão descontínua. Além do Pareamento, é o método mais indicado para a aplicação do modelo aqui proposto. No entanto, deve-se tomar cuidado para excluir aqueles muito longe da linha de corte.

³⁰ Métodos experimentais são aqueles em que há seleção verdadeiramente aleatória entre os grupos de controle e de tratamento. Nos quase-experimentais, essa seleção não é aleatória, mas estatística e artificialmente implementada. A regressão descontínua e o pareamento são exemplos de quase-experimentos.

Um dos métodos não experimentais mais empregados em análises empíricas de causa e efeito é o método da diferença em diferenças. Um dos trabalhos pioneiros na utilização dessa ferramenta é o de Ashenfelter e Card (1985), que busca analisar os efeitos de um programa de treinamento do governo norte-americano sobre os salários dos participantes, em comparação com trabalhadores que não participaram do treinamento.

(SACCARO, 2016, p. 38)

O estimador do método da diferença em diferenças pode ser representado por (SACCARO, 2016):

Equação 9 – Modelo Matemático das Diferença em Diferenças

$$\beta_{DD} = \{E [Y_i|T=1, t=1] - E [Y_i|T=1, t=0]\} - \{E [Y_i|T=0, t=1] - E [Y_i|T=0, t=0]\}$$

Em que “ β_{DD} ” é o efeito causal procurado, “T” é a variável binária de tratamento, que assume valor igual a um caso o indivíduo seja tratado e valor igual a zero, caso contrário. “t” representa o momento e “E(Y)” é o efeito medido do indivíduo Y_i (ANGRIST; PISCHKE, 2009).

Ocorre que Angrist & Pischke (2009, pg. 175) afirmam que é possível o uso de regressão linear para estimar equações como a acima, com o uso do seguinte modelo:

Equação 10 – Regressão linear em Diferença em Diferenças

$$Y_{ist} = \alpha + \gamma J_s + \delta d_t + \beta (J_s \cdot d_t) + \varepsilon_{ist}$$

Em que “s” representa um subgrupo, “t” o período, “i” o indivíduo, γ seria o efeito do indicador J , dependente do grupo e independente do tempo e δ o efeito do indicador d , dependente do tempo e independente do grupo. “ ε ” mede o erro e $\beta (J_s \cdot d_t) = D_{st}$ são os efeitos dependentes do tempo e do subgrupo (ANGRIST; PISCHKE, 2009, p. 175).

Comparando com a primeira equação, temos que:

- $\alpha = E [Y_i|T=1, t=0]$ (estado inicial do grupo de tratamento);
- $\gamma = E [Y_i|T=0, t=0] - \alpha$
- $\delta = E [Y_i|T=1, t=1] - \alpha$
- $\beta = \{E [Y_i|T=1, t=1] - E [Y_i|T=1, t=0]\} - \{E [Y_i|T=0, t=1] - E [Y_i|T=0, t=0]\}$

Ou seja, β nos dá a estimativa da inferência causal (estimativa do impacto) no método da diferença em diferenças (ANGRIST; PISCHKE, 2009). O uso da regressão linear traz vantagens ao modelo, pois mais variáveis independentes podem ser adicionadas bem como medidas estatísticas retiradas (pode-se medir a correlação entre as variáveis dependentes e independentes, dando estimativa da inferência realizada).

O método foi utilizado por Resende et al., (2020), que usou do método quase-

experimental de diferença em diferenças por meio de uma regressão linear para inferir se há causalidade entre ensino integral e melhoria nas notas de português, matemática, redução do trabalho infantil ou aumento da taxa de emprego das mães. Resende et al. (2020) encontraram que não se pode inferir qualquer relação entre estar matriculado em regime integral e quaisquer desses indicadores. No estudo citado, o método quase-experimental de diferenças em diferenças com pareamento foi recomendado como um dos melhores modelos de avaliação de impacto (RESENDE et al., 2020).

O método das diferenças em diferenças é uma abordagem quase-experimental que busca inferir a causalidade de um fenômeno. Rezende (2017) indica que inferências podem ser geradas com confiabilidade, validade, e replicabilidade, atendidos os preceitos e princípios básicos e que os desenhos de pesquisa têm primazia sobre a análise, indicando que a forma como a análise é realizada é primordial. Também indica que, na falta de possibilidade de se realizar um experimento, “que estes desenhos (dos quase-experimentos) enfrentam obstáculos e ameaças similares aos experimentos no que se refere à geração de inferências causais, o que garante que eles sejam considerados como “se fossem”, em alguma medida, experimentos” (REZENDE, 2017).

Finalmente, ele ainda indica que os quase-experimentos é um dos métodos que estão se traduzindo em artigos de alto impacto, num Pluralismo Inferencial, em que coexistem alternativas válidas para resolver o problema fundamental da inferência causal.

13.3. Modelo Econométrico – Regressão Linear

Levando tudo isso em consideração, para calcular o impacto da pesquisa, se torna necessário a comparação do indicador a seu contrafactual pelo método quase-experimental. Para isso, a seguinte regressão linear é proposta:

Equação 11 – Modelo Econométrico para o Cálculo do Impacto.

$$Indicador_{atual,i} = C + I * Indicador_{anterior,i} + P * Financiamento_i + \varepsilon_i$$

Fonte: Elaborado pelo autor.

Na Equação acima, “i” representa cada indivíduo, “C” representa uma constante e seria a progressão do indicador que independeria de qualquer financiamento e/ou produtividade anterior e constante no tempo (equivalente a α na equação anterior). Sua justificativa se dá no fato de que pesquisadores possuem, normalmente, uma produção constante. Outra justificativa que se dá para a constante é que quando a distribuição dos erros não é normal, recomenda-se o uso da constante no modelo (PINO, 2014).

“ T ” representa a medida em que um conhecimento anterior facilita o conhecimento posterior. Todeschini *et al.* (2016) demonstram que a experiência anterior serve como facilitador para a produção de novos conhecimentos. Assim, quem tem mais artigos publicados, por exemplo, tem a tendência de publicar mais artigos. A constante “ I ” busca resgatar esse fato e se equivale a γ na equação anterior. Já ε_i é o erro para cada indivíduo, de modo que se busca minimizar $\sum_i \varepsilon_i^2$ na regressão linear.

Como este modelo busca analisar apenas dois pontos no tempo, não se trata de modelo temporal e d_t pode ser retirado da equação (ANGRIST; PISCHKE, 2009, p. 193). Finalmente, a constante P busca resgatar o impacto do subsídio à P&D, se equivalendo a β na equação anterior, representando, portanto, o resultado da aplicação do método das diferenças em diferenças.

Como considerações, o modelo assume que o indicador no ano atual se correlaciona principalmente com o mesmo indicador no ano anterior, o financiamento realizado e a constante. O modelo deriva de estudos cientométricos e pressupõe *ceteris paribus*³¹ nos demais indicadores e condições. A suposição pode ser aparentemente forte, mas o modelo das diferenças em diferenças já corrige para fatores observados e não observados, pressupondo apenas o mesmo efeito no tempo entre os grupos de tratamento e de controle.

Ou seja, o modelo já corrige para eventuais diferenças causadas por outros indicadores, inclusive, por exemplo, financiamentos externos, inserção acadêmica e formação prévia. Dessa forma, a inclusão de um número maior de indicadores nas variáveis independentes teoricamente pouco ajudaria a melhorar a descritibilidade do modelo e prejudicaria o cálculo do impacto, por causar previsão a menor do impacto causado. O impacto a menor medido deriva do fato de que quanto mais variáveis independentes são adicionadas, necessariamente maior será o a previsibilidade do modelo e menor será o coeficiente de cada uma dessas, independentemente se fizerem sentido ao modelo.

Não se trata de regressão temporal, pois apenas dois anos estão sendo estudados. Dessa forma, o modelo não pressupõe ou assume eventual linearidade da curva, mas apenas que a eventual tendência não linear de ambos os grupos possa ser traduzida pela mesma média da inclinação linear da regressão realizada. Ou seja, assume que a curva no tempo do grupo de tratamento seja similar da do grupo de controle. No entanto, por não se tratar de regressão temporal, nada pode ser inferido sobre a forma da curva.

Ainda, o modelo não é adequado para ser usado como preditivo. Em outras palavras,

³¹ Todas as demais variáveis mantidas constantes, ou seja, que os outros indicadores serão constantes ao longo dos anos.

não se pode usar o modelo para buscar inferir quantos artigos determinado pesquisador terá publicado em determinado ano, ou mesmo para procurar inferir eventual impacto de uma política pública ainda não implementada. Além disso, não há medida individualizada do erro, de modo que o modelo não pode ser usado para inferir quaisquer informações de um único determinado pesquisador.

13.4. Conclusão

Este estudo apresentou proposta de modelo econométrico para inferir a causalidade entre um indicador do Currículo Lattes em períodos diferentes do tempo, dando uma medida do impacto de uma política pública de investimento em ciência. Para isso, o Currículo Lattes pode ser usado para extrair, por meio de Data Mining, ao menos 140 diferentes indicadores que apresentam dados temporais, tornando possível medir o impacto de uma política pública nas dimensões econômicas, sociais, culturais e ambientais.

Para que o modelo econométrico seja usado é necessária a realização de uma regressão linear ordinária nos grupos de tratamento e de controle. Os grupos de controle devem ser escolhidos, preferencialmente, pela seguinte ordem: seleção aleatória, o uso de variáveis instrumentais, regressão descontínua ou pareamento.

O método proposto é escalável, transparente, permitem comparação entre avaliações diferentes e adequado a uso frequente e longitudinal. Traz medida quantitativa do impacto econômico, social, ambiental e cultural e permite o cálculo da eficácia do instrumento analisado. No entanto, não permite ser preditivo, explicativo, nem ir além dos indicadores analisados. Portanto, é ineficaz para explicar os motivos do impacto medido ou para buscar impactos não representados pelos indicadores escolhidos. Por isso, o desenho e os estudos prévios são essenciais para a fundamentação das hipóteses de análise. O modelo apresentado cumpre o objetivo proposto já que permite a avaliação de políticas públicas de investimento em ciência a partir de vários indicadores presentes no Currículo Lattes do CNPq.

14. ANEXO II – ARTIGO: O IMPACTO DE BOLSAS DE DOUTORADO NA PÓS-GRADUAÇÃO BRASILEIRA: UMA ANÁLISE PELO CURRÍCULO LATTES DO CNPQ

The Impact of Doctoral Scholarships on Brazilian Graduate Studies: An Analysis of the CNPq Lattes Curriculum

El impacto de las becas de doctorado en los estudios de posgrado brasileños: un análisis del currículo del CNPq Lattes

Alberto de Campos e Silva

ORCID: <https://orcid.org/0000-0001-6506-1988>

Universidade Federal do Rio Grande do Sul
Programa de Pós-Graduação em Educação em Ciências
Brasília, DF, Brasil

Alexandre Guilherme Motta Sarmiento

ORCID: <https://orcid.org/0000-0002-7201-4466>

Universidade Federal do Rio Grande do Sul
Programa de Pós-Graduação em Educação em Ciências
Brasília, DF, Brasil

Resumo: Este artigo buscou uma medida do impacto do financiamento realizado pela CAPES e pelo CNPq, analisando o Currículo Lattes daqueles que iniciaram seu doutorado no ano de 2014. Para isso, usou-se de análise econométrica quase-experimental, por meio de regressão linear multivariada e o método de diferença em diferenças, chegando-se em análise custo-efetividade. Encontrou-se que mais de 50% dos doutorandos receberam bolsas e que houve impacto, tanto na CAPES quanto no CNPq, em diversos indicadores, incluindo vínculos empregatícios e número de artigos publicados.

Palavras-chave: Avaliação de Impacto, Cientometria, Bolsas de Doutorado, Currículo Lattes, Indicadores Bibliométricos.

Abstract: This article analyzed the Lattes Curriculum of those who started their doctorate in 2014 seeking a measure of the impact of grants provided by CAPES and CNPq. A quasi-experimental econometric analysis was used through linear multivariate regression analysis and the difference-in-differences method, arriving at cost-effectiveness analysis. It was found that more than 50% of doctoral students received scholarships and that there was an impact, both on CAPES and CNPq, on several indicators, including employment relationships and number of published articles.

Keywords: Impact Evaluation, Scientometrics, PhD Scholarships, Lattes Curriculum, Bibliometric Indicators.

Resumen: Este artículo buscó medir el impacto de la financiación proporcionada por CAPES y CNPq, analizando el Currículo Lattes de aquellos que iniciaron su doctorado en 2014. Para ello, se utilizó el análisis econométrico cuasi-experimental, a través del análisis multivariado de regresión lineal y el método de diferencias en diferencias, llegando al análisis de costo-efectividad. Se constató que más del 50% de los doctorandos recibieron becas y que hubo impacto, tanto en la CAPES como en el CNPq, en varios indicadores, incluyendo relaciones laborales y número de artículos publicados.

Palabras clave: Evaluación de Impacto, Cienciometría, Becas de Doctorado, Currículo Lattes, Indicadores Bibliométricos.

14.1. Introdução

Ao formar doutores, as universidades e instituições de pesquisa contribuem diretamente na qualidade da educação no país, produzindo profissionais capazes de desenvolver pesquisas relevantes e inovadoras e contribuindo para o desenvolvimento social e econômico do país. Como exemplo, o nível de ensino de professores da educação básica evoluiu juntamente com o crescimento da Pós-graduação *stricto sensu*, mostrando a importância e a preocupação com formação continuada de professores. Apesar disso, mesmo com ações de fomento se acumulando ao investimento natural nos últimos anos, o sistema educacional carece de maior esforço em sua cadeia de formação para a qualificação de seus profissionais e o respectivo aumento de professores pós-graduados (SILVA; PORCIÚNCULA, 2017).

Nesse sentido, o [Plano Nacional de Educação](#) (PNE) determina diretrizes, metas e estratégias para a política educacional no período de 2014 a 2024 no intuito de promover a formação de mestres e doutores. Entre as [20 metas](#) do Plano, as de nº 13 e 14 envolvem a formação de doutores no país: “elevar a qualidade da educação superior e ampliar a proporção de mestres e doutores do corpo docente em efetivo exercício no conjunto do sistema de educação superior para 75% (setenta e cinco por cento), sendo, do total, no mínimo, 35% (trinta e cinco por cento) doutores” e “elevar gradualmente o número de matrículas na pós-graduação de modo a atingir a titulação anual de 60.000 (sessenta mil) mestres e 25.000 (vinte e cinco mil) doutores” (PNE - Plano Nacional de Educação, n.d.).

De acordo com o MEC, um dos aspectos que influencia a qualidade da educação superior é a titulação do corpo docente, sobretudo em cursos de mestrado e doutorado. Uma das estratégias para o cumprimento da meta relacionada é elevar o padrão de qualidade das universidades, visando à realização de pesquisa institucionalizada de modo articulado a programas de pós-graduação *stricto sensu*. Entre as estratégias estabelecidas no novo PNE, encontram-se expandir o financiamento da pós-graduação *stricto sensu* por meio das agências oficiais de fomento, incentivar a atuação em rede e o fortalecimento de grupos de pesquisa e promover o intercâmbio científico e tecnológico, nacional e internacional, entre as instituições de ensino, pesquisa e extensão (PRESIDÊNCIA DA REPÚBLICA; MINISTÉRIO DA EDUCAÇÃO; SECRETARIA DE ARTICULAÇÃO COM OS SISTEMAS DE ENSINO, 2014).

Assim, é na pós-graduação, principalmente no nível de doutorado, que pesquisa e educação possuem forte área de intersecção. Nesse sentido, as estratégias de fortalecimento da pós-graduação do governo federal do Brasil duas agências de fomento: a CAPES e o CNPq. A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), é uma Fundação do Ministério da Educação (MEC), e tem como missão a expansão e consolidação da pós-graduação stricto sensu (mestrado e doutorado) no Brasil (“História e missão — CAPES”, [s.d.]). Para cumprir seu objetivo, a CAPES concede bolsas de estudo no Brasil visando a formação de doutores. Por outro lado, a missão institucional do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) é a de fomentar a pesquisa científica, tecnológica e de inovação e contribuir com políticas públicas, para fortalecer o desenvolvimento do Brasil (“Apresentação — Conselho Nacional de Desenvolvimento Científico e Tecnológico”, [s.d.]). Dessa forma, o CNPq fornece bolsas de iniciação científica, mestrado e de doutorado, tanto dentro de projetos científicos, quanto como na forma de cotas para programas de pós-graduação, no intuito de fomentar a pesquisa científica. Finalmente, o CNPq, ao diretamente fomentar projetos de pesquisas, indiretamente também contribui para a formação de doutores no Brasil.

No entanto, carecem estudos sobre o impacto do financiamento dessas agências em bolsas de doutorado. Assim, este artigo buscou análise cientométrica dos currículos daqueles que iniciaram seu doutorado com recorte do ano de 2014, comparando indicadores dos anos 2013 e 2019, buscando identificar o impacto do financiamento de bolsas de doutorado pelo CNPq e pela CAPES. A análise efetuada foi quase-experimental³², por meio de regressão linear multivariada³³, incluindo variáveis categóricas³⁴, e o modelo de diferença em diferenças³⁵. Ainda, esse impacto foi comparado ao investimento realizado pelo CNPq, chegando-se em análise custo-efetividade³⁶.

³² Na análise experimental os grupos de tratamento e de controle são escolhidos aleatoriamente. Na quase-experimental, se usa de artifícios estatísticos para a escolha desses grupos, de modo a reduzir o viés.

³³ Regressão com uma variável dependente e duas ou mais independentes. Aqui, a variável dependente é o indicador em 2019, e as múltiplas independentes são o indicador em 2014 e a concessão ou não de bolsas.

³⁴ Variável não numérica, não contínua, que representa uma categoria. Por exemplo, com ou não bolsas da CAPES.

³⁵ Será detalhado adiante, mas é método que compara os grupos de tratamento e o de controle em dois períodos diferentes para encontrar o impacto da política pública aplicada.

³⁶ Cálculo de quanto custa cada progresso no indicador. Por exemplo, cada artigo publicado custou x reais.

14.2. Metodologia

No contexto da ciência aberta, todo o código fonte usado pode ser encontrado no GitHub https://github.com/AlbertoCamposSilva/CV_Lattes_Analyzer. Lá também podem ser encontradas as tabelas necessárias para rodar os códigos aqui indicados, bem como planilhas com os resultados completos das diversas regressões realizadas, retirados todos os identificadores pessoais.

Para a escolha dos dados a serem analisados considerou-se que o objetivo desta pesquisa é encontrar o impacto da concessão de bolsas de doutorado. Como o tempo de doutoramento pode ser de até cinco anos, considerando que a pandemia de 2020 pode causar efeitos imprevisíveis, filtrou-se todos os pesquisadores que iniciaram seu primeiro doutorado no ano de 2014³⁷, comparando o currículo de 2013 (antes do início do mesmo) com o de 2019 (após a sua provável conclusão). A base de dados utilizada para a análise cientométrica foi o Currículo Lattes, baixados através do Extrator Lattes³⁸.

O Currículo Lattes (Lattes) foi criado em 1990 pelo Governo Federal como forma de padronização dos currículos acadêmicos. O Lattes é obrigatório para os pesquisadores em instituições públicas no Brasil, sendo requisito para a concessão de bolsas e progressões na carreira. Assim, o Lattes se consolidou como fonte de dados para análise bibliométrica (PERLIN; SANTOS, 2015). Para esta pesquisa, foram identificados 7.427.277³⁹ currículos Lattes. Do total dos currículos, apenas 2.652.105 foram atualizados a partir de 01/01/2020, seleção necessária uma vez que se pretende comparar o ano de 2014 como de 2019.

Cada um desses currículos foi baixado e tratado de modo a se retirar, do mesmo, indicadores⁴⁰. Um indicador é aqui referido como um dado, presente no Lattes, com dado temporal definível. Como exemplo, vamos supor que uma determinada pessoa com Lattes atualizado após o ano de 2020 tenha iniciado seu doutorado em educação pela UFRGS em 2014 e concluído o mesmo em 2018. Desses dados são retirados dois indicadores: Início do doutorado (2014) e conclusão do doutorado (2018), sendo ignoradas todas as outras informações referentes ao mesmo, por

³⁷ Código Fonte: 16.2 - Importa ao BD todos os CVs com nível 4 (doutorado), 16.2 Atualiza todos os indicadores de doutorandos 2014

³⁸ <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>

³⁹ Números Ids Lattes diferentes constantes do arquivo em <http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>.

⁴⁰ Código fonte: Função “atualiza()” em Indicadores.py e 16.2 - Importa ao BD todos os CVs.ipynb

ausência de ano relacionado. Assim, dos 2.649.498⁴¹ currículos em que foram retirados indicadores, encontrou-se um total de 21.250⁴² currículos de pessoas que iniciaram o doutorado em 2014 e o atualizaram em 01/01/2020 ou depois.

Nota-se que os dados apresentados trazem uma limitação na pesquisa, que é a não indicação do impacto do financiamento realizado nas chances de titulação. Isso porque os dados disponíveis poderiam trazer viés, já que há quem iniciou o doutorado em 2014, mas o fato não está indicado em seu currículo (ou porque esse não foi atualizado, ou porque o doutoramento inconcluso foi excluído do mesmo)⁴³. Para evitar outro tipo de viés (quem ainda pretende se formar mas ainda não concluiu o curso), não foram retirados os que não se formaram.

Os currículos foram então separados em quatro distintos grupos: aqueles que foram financiados pela CAPES; os com bolsas do CNPq; os com bolsa de ambas as agências; e os que não mencionam qualquer bolsa recebida⁴⁴. O dado sobre o financiamento ou não pela CAPES foi retirado do Lattes, em campo específico em que consta a CAPES como agência financiadora do doutorado, de modo que a exatidão do dado depende de o ex-bolsista ter preenchido corretamente o Lattes⁴⁵. O financiamento pelo CNPq foi retirado do portal de dados abertos⁴⁶, de modo que se presume ser dado exato e preciso, salvo eventuais homônimos. Deve-se mencionar que, mesmo após tentativas de se contatar a CAPES, não foi possível encontrar dados semelhantes aos fornecidos pelo CNPq.

O estudo então traça os perfis comparativos entre esses grupos, por meio de regressão linear multivariada com variáveis categóricas:

Equação 12 – Medida de um Impacto por meio de Regressão multivariada categórica

$$Indicador_{2019} = \alpha + \beta Indicador_{2013} + \gamma CAPES + \delta CNPq$$

⁴¹ A diferença deste número com o anterior é devido a erros de importação, todos com origem no Extrator Lattes. Infere-se que sejam currículos cancelados ou excluídos.

⁴² Já excluídos os Ids com erro de importação. SQL: Lista IDs doutorado 2014 para pareamento.sql

⁴³ Pesquisa informal identificou que muitos doutorandos simplesmente tiraram o doutorado inconcluso do currículo, ao invés de o identificar como cancelado, impedindo estudos comparativos entre financiamento e seus efeitos nas chances de conclusão do mesmo.

⁴⁴ Lista das variáveis categóricas inclusas no estudo: CAPES, CNPq, Ambos e Nenhum.

⁴⁵ Código Fonte: 16.6 Cria a tabela financiamentos com lista dos IDs que se doutoraram em 2014 e foram ou não financiados.

⁴⁶ http://dadosabertos.cnpq.br/pt_BR/

Na equação acima, $Indicador_{2013}$ e $Indicador_{2019}$ representam a soma de todas as entradas desse indicador até os anos de referência (2013 e 2019), como, por exemplo, o número total de artigos publicados por um determinado pesquisador até os anos 2013 e 2019, respectivamente. CAPES é uma variável categórica igual a 1, caso o pesquisador tenha recebido bolsas dessa agência, ou 0 caso não tenha, sendo a mesma metodologia aplicada à variável CNPq (sendo os casos de duplo financiamento descartados). Dessa forma, α representa um aumento constante do indicador no tempo, β o aumento temporal desse indicador em relação a seu estado inicial e aqueles com duplo financiamento foram retirados da pesquisa.

A metodologia é aquela recomendada pelo Banco Mundial em seu livro Avaliação de Impacto na Prática (GERTLER et al., 2018). Como não é possível método experimental para medir o impacto⁴⁷, utiliza-se a segunda melhor opção possível, que é a metodologia quase-experimental. Essa metodologia busca estimar os efeitos do tempo, bem como todos os outros desconhecidos, comparando o estado do Lattes no ano de 2019 com aquele do ano de 2013 de quem não foi financiado, utilizando-se esse como grupo de controle.

Assim, a diferença entre os estados dos currículos do grupo de controle com o estado dos currículos nos grupos de tratamento busca eliminar do grupo de tratamento eventual viés decorrente de efeitos externos e dando estimativa do impacto do investimento realizado, o que consiste na metodologia quase-experimental da diferença em diferenças. Neste caso, essa metodologia é aplicada como modelo da regressão linear já exposta na Equação 12. Assim, $\alpha + \beta Indicador_{2013}$ representam os efeitos no grupo de controle, enquanto $\gamma CAPES + \delta CNPq$ representam uma medida do impacto das bolsas concedidas em cada um dos grupos de tratamento.

A fragilidade da metodologia está na possibilidade de viés no grupo de controle (chamado viés de seleção), ou seja, se a inclinação da reta de progressão no tempo do grupo de controle se diferir da do grupo de tratamento. Um exemplo seria a seleção para o recebimento de bolsas daqueles candidatos com a maior possibilidade de publicar artigos. Nesse caso, a inclinação da reta do grupo de controle seria menor, com menos chances de publicar artigos, que o grupo de tratamento, mesmo sem concessão de bolsas.

⁴⁷ O método experimental significaria sortear quem ganharia bolsa de mestrado e quem não ganharia, algo impraticável por motivos práticos, políticos e legais.

Observa-se que, de acordo com (GERTLER et al., 2018), mesmo presente eventual viés, a metodologia usada é a indicada e a capaz de dar os resultados mais próximos da realidade, tendendo a dar resultado subestimado, uma vez que usa de rigor estatístico para desconsiderar indicadores que não apresentaram o nível de confiança exigido.

Para buscar diminuir, ou medir, eventuais efeitos do viés acima mencionado, também foi realizada a metodologia do pareamento. De acordo com (GERTLER et al., 2018), uma das formas de se reduzir o viés é buscar alguém com o mesmo perfil (par) e realizar a metodologia acima exposta considerando esse par como o grupo de controle.

Para realizar o pareamento, foi considerada a seleção já indicada (início do doutorado em 2014, atualização do currículo a partir de 2020) e depois o gênero. A seleção pelo gênero se deu pelos diversos estudos existentes indicativos de que mulheres possuem certa desvantagem na ciência em relação aos homens (AREAS et al., 2020). Com essas considerações, todos os indicadores do Lattes foram separados em grupos (listados abaixo) e calculou-se os erros quadrados médios de cada um desses grupos de indicadores, selecionando-se, para cada ex-bolsista do CNPq, o não bolsista, cuja soma dos erros ao quadrado era a menor⁴⁸. A amostra foi sem reposição, ou seja, cada doutorando selecionado não poderia ser novamente selecionado, uma vez que a amostragem com reposição verificou forte tendência a que um doutorando fosse selecionado como par de diversos outros. A regressão anteriormente indicada também foi alterada para considerar o valor financiado pelo CNPq. Assim, a equação usada para a regressão foi:

Equação 13 – Medida de Impacto por meio de Regressão multivariada categórica para o CNPq:

$$Indicador_{2019} = \alpha + \beta Indicador_{2013} + \theta Financiamento_{CNPq}$$

Em que $Indicador_{2019}$, α , β e $Indicador_{2013}$ mantêm seus significados originais (valor dos indicadores nos respectivos anos, constante e a variável que dá o efeito do indicador em 2013 no indicador em 2019). Mas, $Financiamento_{CNPq}$ agora mede o valor financiado⁴⁹ e θ o impacto desse financiamento. Assim, mede-se mais precisamente o impacto da concessão de

⁴⁸ Código Fonte: 16.7 Pareamento de Doutorados.ipynb

⁴⁹ Para melhor precisão e cálculo instantâneo do impacto, a variável sofreu transformada, dividindo-se o montante financiado pela constante do valor total financiado pelo CNPq. A transformada não afeta o resultado final e faz com que a regressão resulte diretamente no impacto total.

bolsas de doutorado pelo CNPq, algo não possível para a CAPES devido a já mencionada falta de dados.

Foi então realizada regressão para cada um dos 140 indicadores, sendo descartadas aquelas com impacto em menos de 1% da população, as não significativas ($P|t| > 0,01$) e as com $R^2_{\text{ajustado}} < 30\%$, método repetido ao longo deste estudo.

R^2_{ajustado} é uma medida que busca indicar o quanto da variância dos dados pode ser explicada pela fórmula encontrada e, portanto, o quanto a fórmula consegue “explicar o mundo real”. R^2_{ajustado} inferior a 0,3 indica, mesmo se houver correlação estatisticamente significativa, essa é “fraca” e não consegue descrever corretamente a variação dos dados (GRUS, 2019). Assim, a representatividade é:

- Fraca: $R^2_{\text{ajustado}} < 0,3$
- Média: $0,3 \leq R^2_{\text{ajustado}} \leq 0,7$
- Forte: $R^2_{\text{ajustado}} > 0,7$

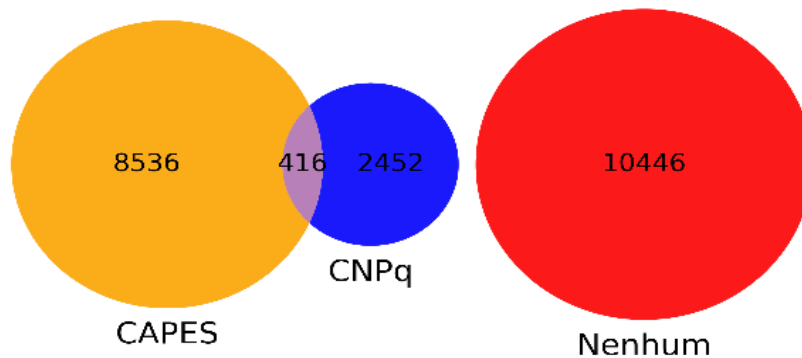
14.3. Análise dos dados

14.3.1. Distribuição dos Financiamentos

A primeira análise realizada é em relação ao diagrama de Venn da Figura 27⁵⁰. Nota-se que 52% daqueles que iniciaram seu doutorado receberam bolsas de estudo. O Número está de acordo com Fundação Nacional de Ciência dos Estados Unidos da América (NSF, National Science Foundation), cuja análise do senso realizado permite inferir que aproximadamente 53% dos doutorandos nas áreas de ciência, engenharia e saúde receberam apoio financeiro na forma de subsídios, bolsas de estudo ou assistência durante seus estudos de pós-graduação. Os dados, tanto nacionais quanto internacionais, demonstram a importância do financiamento a estudantes de bolsas de doutorado. (“Doctorate Recipients from U.S. Universities: 2021 | NSF - National Science Foundation”, [s.d.]

⁵⁰ Código Fonte: 15.5 Diagrama de Venn

Figura 27 - Distribuição da concessão de bolsas de doutorado daqueles que iniciaram seu doutorado em 2014.



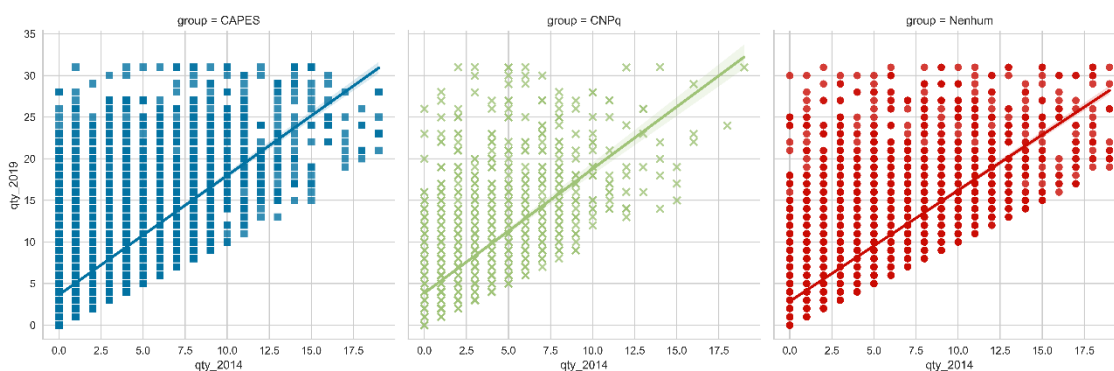
Fonte: autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

A segunda conclusão a que se chega do mesmo diagrama é que o CNPq financiou apenas 13% dos doutorandos, enquanto a CAPES concedeu bolsas a 41%. Dessa forma, nota-se que a presença da CAPES nesse assunto é maior que a do CNPq.

14.3.2. Distribuição dos Dados

Um dos pressupostos mais usuais em ciência de dados e também um dos que menos são verdadeiros é o de que os dados (ou a distribuição dos erros em caso de regressão linear) estejam normalmente distribuídos. As implicações da não normalidade são diversas e podem depender de cada caso (GRUS, 2019), tornando o estudo necessário. Quanto à normalidade dos dados, usou-se o teste de Shapiro Wilk, em que se encontrou a não normalidade (p -valor menor que 0,01). A distribuição do indicador Artigos Publicados consta da Figura 28.

Figura 28 – Distribuição do indicador Artigos Publicados de acordo com a concessão de bolsas



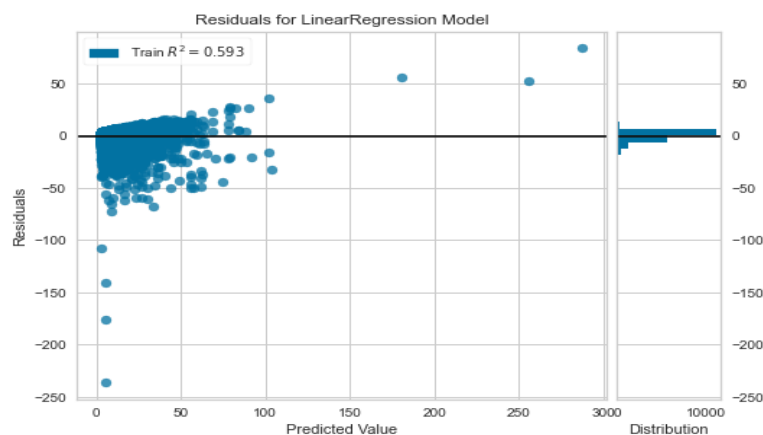
Fonte: Autoria própria (2023).

Para verificação matemática da distribuição dos erros nas regressões realizadas, verificou-se, nas regressões realizadas, que o “Prob(Omnibus)”, que é o teste que busca verificar a hipótese

nula de que a distribuição dos erros é normal, foi normalmente inferior a 0.000. Ou seja, temos mais de 99,9% de chance de que os erros não estão normalmente distribuídos.

Tentou-se então várias transformadas para a busca de uma distribuição normal, sendo o melhor resultado obtido com as seguintes transformadas: remoção dos *outliers* (pontos acima de quatro vezes o desvio padrão), aplicação da fórmula $\sqrt{\log(\text{indicador})}$ nas somas tanto de 2019 quanto na de 2013 e remoção de todos os pontos NaN (not a number). Mesmo com essas transformadas, encontrou-se que a distribuição dos erros não era próxima à curva normal (Prob(Omnibus): 0.000) e verificou-se que R^2 diminuiu significativamente. Dessa forma, optou-se por não realizar nenhuma transformada ou exclusão de *outliers*.

Figura 29 – Distribuição dos Erros da Regressão Linear do Indicador Artigos Publicados



Fonte: Autoria própria (2023)

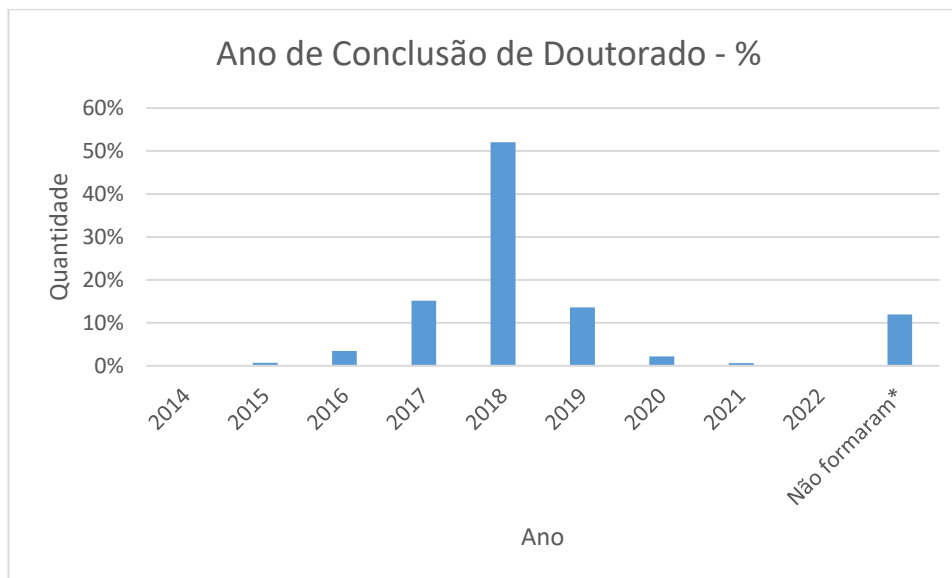
A principal consequência da não normalidade dos erros está na capacidade preditiva da equação, mais sensibilidade a *outliers*, redução da precisão do teste de significância e da precisão em grupos de tamanhos distintos. Mesmo assim, como há muitos dados disponíveis, o teorema do limite central indica a normalidade assintótica e a correta aplicação da ferramenta estatística (BERNIER; FENG; AZAKAWA, 2011 *apud* ALBERTO PINO; PAULO, [s. d.]), principalmente porque não se buscam modelos preditivos. Além disso, observa-se que o uso da constante α é recomendada devido à não normalidade dos erros (PINO, 2014).

A regressão linear é, portanto, ferramenta indicada para que se chegue ao mínimo dos erros quadráticos e a melhor ferramenta para dar o impacto médio da concessão de bolsas.

14.3.3. Ano de Conclusão do Doutorado

Analisando a relação do tempo de conclusão do doutorado, temos a seguinte distribuição:

Figura 30- Ano de conclusão do doutorado iniciado em 2014 (%).

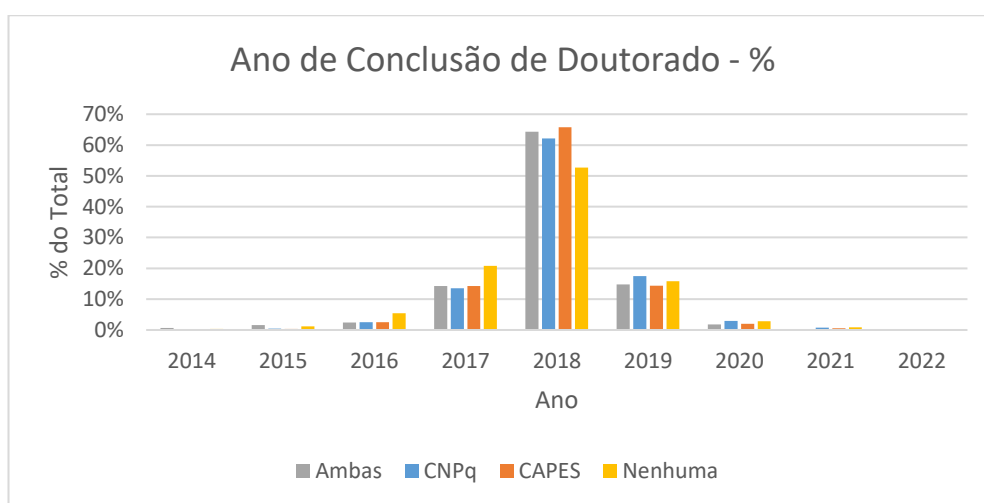
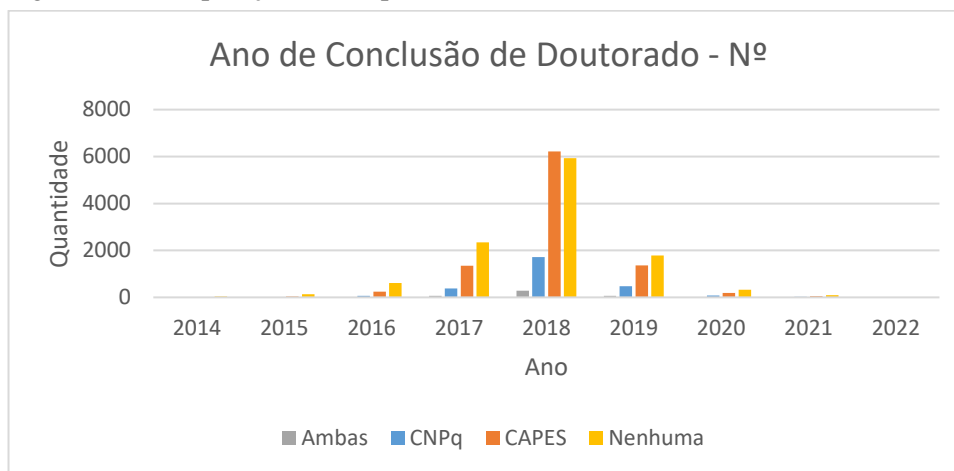


Fonte: autoria própria⁵¹.

A Figura 30 apresenta que 59% das pessoas que iniciaram seu doutorado em 2014 o concluíram em 4 anos e que 15% da população não concluiu o doutorado até o ano de 2019, sendo que 12% não o concluíram até o ano de 2022. No entanto, há a ressalva de que esse dado de 12% pode não representar a realidade, uma vez que o Lattes é autodeclaratório e uma vez que se contatou casos em que a falha na obtenção do título implicou na retirada total, do Lattes, das informações de eventual doutorado não concluído.

⁵¹ Código Fonte: Código: Ano de Conclusão de Doutorado por Financiamento ou Não.sql

Figura 31 - Comparação do tempo de doutoramento - com e sem bolsas de estudo.



Fonte: Autoria própria.

Uma questão que pode ser levantada é se a concessão de bolsas afeta esse tempo de conclusão em alguma forma. Para o levantamento, os doutorandos foram separados em dois grupos distintos: com bolsas e sem bolsas. Assim, foi realizada regressão linear com o seguinte modelo:

Equação 14 - Financiamento influencia no tempo de doutoramento?

$$\gamma_i = \alpha + \beta X_i$$

Na equação acima, γ_i é a subtração do ano de conclusão do doutorado com o ano de início do mesmo (aqui, 2014); α é a constante e representa o tempo médio para que se conclua o doutorado; X_i é variável categórica de valor 1, caso o doutorando tenha recebido bolsa de estudos, ou zero, caso não tenha; de modo que β mede a influência, em anos, do financiamento no tempo de doutoramento. Encontrou-se que:

- $\alpha = 4$ (tempo médio de doutoramento), com $P > |t| = 0,000$ (significativo);
- $\beta = 0.086$ (ser financiado aumenta em um mês o tempo de doutoramento), com $P > |t| = 0,000$ (significativo);
- $R^2_{ajust.} = 0.002$ (ser financiado influencia apenas em 0,2% no tempo de doutoramento); e
- Prob (F-statistic): $3.83e-12$ (a regressão é estatisticamente relevante).

Conclui-se que, embora haja correlação entre a concessão de bolsas e o aumento em um mês do tempo para a titulação, tanto o valor de β quanto o de $R^2_{ajust.}$ são o suficientemente pequenos para podermos concluir que não há impacto da concessão de bolsas no tempo necessário para concluir o doutorado. Vale mencionar que esse mesmo raciocínio será adotado em outras análises abaixo.

Também pode-se concluir que o tempo médio de doutoramento é de cerca de quatro anos e que a janela de cinco anos inclui ao menos 92% dos que se doutoraram, de forma a ser uma janela temporal estatisticamente conclusiva.

14.3.4. Impacto do Financiamento

Foi montada tabela com os Ids de todos os currículos que tiveram o início de seu primeiro doutorado em 2014, resultando em 21.250 linhas⁵². Tabela com os indicadores foi gerada⁵³ e filtrada por indicador⁵⁴, para que então a regressão linear fosse gerada, sendo usado a seguinte fórmula:

Equação 15 – Medida de um Impacto por meio de Regressão multivariada categórica

$$Indicador_{2019} = \alpha + \beta Indicador_{2013} + \gamma CAPES + \delta CNPq$$

Assim, a regressão linear foi realizada em cada um dos 140 indicadores⁵⁵. Dos resultados, retirou-se aqueles cujos Prob (F-statistic) > 0.01 (mais de 1% de chance de que a regressão não é estatisticamente relevante), os com impacto pequeno (menor que 212), e os indicadores cujos $R^2_{ajustado} < 0,3$. Assim, restaram os seguintes indicadores:

⁵² Tabela: ids_doutorandos_inicio_em_2014. SQL: Cria tabela com listan de IDs que iniciaram o doutorado em 2014.sql

⁵³ Código Fonte: 16.2 - Importa ao BD todos os CVs

⁵⁴ SQL: Cria tabela de indicadores para doutorado.sql

⁵⁵ Código Fonte: 16.7 Fazendo Regressão CAPES, CNPQ, Nenhum

Tabela 8 – Regressão Linear com Medida de um Impacto de Concessão de Bolsas de Doutorado

Indicador	Nº	R ² ajustado	α	β	γ	δ	$\frac{\gamma + \delta}{2}$
Artigo Publicado	19173	46%	3,6	1,58	1,39	1,81	1,60
Início de Vínculo Empregatício ou Funcional	20671	67%	1,66	1,29	1,46	1,66	1,56
Início de Atividade de Ensino	9463	72%	2,08	1,24	1,41	1,33	1,37
Fim de Atividade de Ensino	8131	73%	2,21	1,26	1,11	1,09	1,10
Produção de Carta, Mapa ou Similar	252	71%	0	1,22	1,96	0	0,98
Trabalho em Eventos	18283	74%	3,95	1,35	0,95	0,87	0,91
Detalhamento do Trabalho	18248	74%	3,9	1,35	0,91	0,79	0,85
Fim de Vínculo Empregatício ou Funcional	20025	73%	1,5	1,22	0,87	0,74	0,81
Ministério de Curso de Curta Duração	11061	75%	1,03	1,31	0,43	0,52	0,48
Produção de Trabalho Técnico	7032	55%	3,33	1,2	0,77	0	0,39
Conclusão de Curso de Curta Duração	16545	77%	1,63	1,17	0,26	0,36	0,31
Início de Curso de Curta Duração	16568	77%	1,64	1,17	0,26	0,36	0,31
Fim de Atividade de Pesquisa e Desenvolvimento	3505	66%	0,75	0,93	0,17	0,28	0,23
Apresentação de Trabalho Técnico	18894	72%	3,82	1,38	0,42	0	0,21
Início de Estágio	4132	83%	0,23	1	0,21	0,17	0,19
Fim de Estágio	3968	83%	0,23	1	0,21	0,16	0,19
Início de Atividade de Pesquisa e Desenvolvimento	7093	61%	0,83	0,96	0,13	0,15	0,14
Participação em Congresso	18769	79%	1,86	1,22	0	0,24	0,12
Prêmio ou Título Recebido	11599	67%	1,24	1,17	0,13	0	0,07
Participação em Simpósio	14864	74%	0,98	1,13	0,09	0	0,05
Início de Participação em Projeto	16636	70%	1,6	1,23	-0,12	0	- 0,06
Início de Projeto de Pesquisa	16537	61%	2,01	1,3	-0,18	0	- 0,09
Publicação de Capítulo de Livro	11476	43%	1,95	1,42	0	- 0,27	- 0,14
Artigo Aceito para Publicação	3890	38%	0,6	1,03	-0,12	- 0,15	- 0,14
Fim de Projeto de Pesquisa	14815	57%	1,86	1,24	-0,17	- 0,24	- 0,21
Outras Orientações Concluídas	12780	71%	7,53	1,28	-1,33	- 1,92	- 1,63
Participação em Banca de Graduação	13381	72%	7,19	1,24	-1,19	- 2,16	- 1,68

Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Com base na tabela acima, podemos agrupar as atividades com “início” e “fim”, subtraindo o impacto encontrado em “início” daquele que indica o “fim” para se chegar ao quantitativo de uma “Nova” entrada, bem como calcular o impacto. Assim, a seguinte tabela pode ser construída⁵⁶:

Tabela 9 –Impacto de Concessão de Bolsas de Doutorado pelo CNPq e pela CAPES.

Indicador	Nº CAPES	Impacto CAPES	Nº CNPq	Impacto CNPq	Impacto Total
Artigo Publicado	7833	10856	2297	4163	15.019
Trabalho em Eventos	7493	7084	2122	1844	8.927
Detalhamento do Trabalho	7476	6790	2118	1683	8.473
Novo Vínculo Empregatício ou Funcional	142	4951	57	2247	7.198
Apresentação de Trabalho Técnico	7653	3251	2183	0	3.251
Conclusão de Curso de Curta Duração	6770	1764	1851	671	2.435
Ministério de Curso de Curta Duração	4243	1807	1103	577	2.384
Produção de Trabalho Técnico	2675	2060	669	0	2.060
Nova Atividade de Ensino	404	1454	117	353	1.807
Prêmio ou Título Recebido	4521	586	1416	0	586
Participação em Simpósio	6098	535	1754	0	535
Participação em Congresso	7609	0	2210	531	531
Novo Projeto de Pesquisa	685	-176	186	428	253
Produção de Carta, Mapa ou Similar	117	229	26	0	229
Nova Atividade de Pesquisa e Desenvolvimento	1341	125	330	-6	119
Novo Estágio	60	1	11	7	8
Artigo Aceito para Publicação	1561	-192	421	-65	-257
Publicação de Capítulo de Livro	4432	0	1156	-312	-312
Outras Orientações Concluídas	4721	-6265	1233	-2372	-8.637
Participação em Banca de Graduação	5307	-6330	1325	-2856	-9.185

Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Dos dados encontrados acima, analisaremos com mais cuidado apenas dois exemplos, de modo que a análise efetuada possa ser facilmente estendida aos outros casos.

⁵⁶ Tabela: Impacto dos Doutorandos 2014 com CAPES e CNPq.xlsx

14.3.5. Artigos Publicados

O $R^2_{ajustado}$ do indicador relacionado a Artigos Publicados é de 46%, ou seja, a fórmula aplicada consegue explicar cerca de 46% da variação do quantitativo de artigos publicados até 2019. Como a estatística é significativa (o maior valor de $P > |t|$ foi de 4·10⁻³⁰, em δ), temos a ida (forte correlação estatística) e a volta (o modelo adotado descreve mediantemente os dados observados), de forma que podemos inferir que houve impacto. Podemos assim interpretar os números:

- α : para se doutorar, o pesquisador publica em média 3,6 artigos.
- β : já ter publicado antes de iniciar seu doutorado se correlaciona com mais 1,6 artigos publicados por artigo inicial e até o ano de 2019.
- γ : ter recebido bolsas da CAPES se correlaciona com mais 1,4 artigos publicados até o ano de 2019.
- δ : ter recebido bolsas do CNPq se correlaciona com mais 1,8 artigos publicados até o ano de 2019.

A amostra contempla apenas 19.173 dos 21.250 doutorandos analisados, o que indica que 1.714 doutorandos não publicaram artigos até o ano de 2019, não os incluíram em seu Lattes ou foram excluídos por terem recebidos bolsas de ambas as agências (416 casos).

14.3.6. Vínculos

Há 15.948 doutorandos que iniciaram algum tipo de vínculo (com instituições) até o ano de 2019 e 8.750 que encerraram qualquer tipo de vínculo. Examinados em separados, esses números são pouco significativos, assim como qualquer indicador que tenha “Início” e “Fim” em seus nomes. Mas, examinados em conjunto, podemos verificar que houve 7.198 novos vínculos firmados devido ao doutorado.

Vínculo é uma entrada no Lattes, de forma que qualquer alteração desse campo no Lattes é contabilizada. Ou seja, os números não indicam 7.198 novos empregos, mas alterações nos vínculos, que podem ser, por exemplo, uma mudança de lotação, uma promoção, um novo emprego ou mesmo o desemprego.

Como exemplo, há casos de vínculos “bolsista da CAPES”, o que explica parte dos 4.951 novos vínculos devido aos financiamentos pela CAPES. Por outro lado, verificou-se a falta de

atualização de alguns currículos quando do término da bolsa da CAPES, de modo que o bolsista registra o início dessa, mas não seu término.

Finalmente, analisando apenas o grupo de controle:

Equação 16 - Impacto Negativo em Novos Vínculos

$$Indicador_{2019} = \alpha + \beta Indicador_{2013}$$

$$Impacto = \{(1,66 * 20.671) - (1,50 * 20.025)\} = -4.235$$

O impacto negativo de 19% dá indicativo de desistência de vínculos durante o processo de titulação para quem não recebeu bolsas, algo que pode ser melhor analisado.

14.3.7. Pareamento

Para análise do Impacto da concessão de bolsas pelo CNPq, pode-se adotar metodologia em que se escolhe, para cada pessoa tratada, uma pessoa semelhante para compor o grupo de tratamento, em busca de eliminar possíveis vieses de seleção (i.e. quando a política de concessão escolhe o grupo tratado por variáveis capazes de influenciar no resultado) (GERTLER et al., 2018). Dessa forma, foi realizado o processo de pareamento⁵⁷, conforme metodologia indicada pelo livro do banco mundial Avaliação de Impacto na Prática (GERTLER et al., 2018). Assim, para cada doutorando que recebeu financiamento do CNPq, foi escolhido um par da seguinte forma:

- Iniciou seu doutorado em 2014,
- Atualizou seu Lattes a partir de 01/01/2020,
- Do mesmo sexo,
- Com a minimização de erros quadráticos nos seguintes grupos de indicadores:
 - Atividades de Pesquisa e Desenvolvimento,
 - Livros Publicados,
 - Atividades de Ensino,
 - Publicações de Artigos,
 - Vínculos Empregatícios ou Funcionais,
 - Mestrados,

⁵⁷ 16.7 Pareamento de Doutorados

- Patentes e
- Participação em Congressos.

Para a minimização dos erros quadráticos, para cada doutorando financiado pelo CNPq calculou-se a diferença da soma dos indicadores em cada grupo, ao quadrado. Assim, por exemplo, se um doutorando teve 2 Livros Publicados e outro nenhum, o erro de “Livros Publicados” é de $(2-0)^2 = 4$. Somando-se esse erro para todos os outros grupos, temos a soma dos erros quadráticos.

Uma vez realizado o pareamento, pode-se aplicar o método da diferença em diferenças calculando a diferença dos indicadores no grupo de tratamento com os do grupo de controle. Ou seja:

Equação 17 - Diferença em Diferenças

Diferença em Diferenças

$$= [Estado_{atual} - Estado_{inicial}]_{Grupo\ de\ Tratamento} - [Estado_{atual} - Estado_{inicial}]_{Grupo\ de\ Controle}$$

A equação acima, levada para nosso caso em questão, resulta em:

Equação 18 – Cálculo do Impacto pelo método da Diferença em Diferenças

$$Impacto = [Indicador_{2019} - Indicador_{2013}]_{Com\ Bolsas\ do\ CNPq} - [Indicador_{2019} - Indicador_{2013}]_{Sem\ Bolsas\ do\ CNPq}$$

Aplicando a fórmula acima ao Banco de Dados⁵⁸, temos o seguinte impacto⁵⁹, sendo aqui listadas apenas as 10 primeiras e 3 últimas linhas, ordenadas pelo Impacto:

Tabela 10 - Diferença em Diferenças por Pareamento dos Doutorandos 2014 por soma simples.

Indicador	Sem Financiamento			Financiado pelo CNPq			Impacto
	2019	2013	Diferença	2019	2013	Diferença	
Trabalho em Eventos	22451	13012	9439	40688	22857	17831	8392
Artigo Publicado	10446	2991	7455	18442	4165	14277	6822
Início de Vínculo Empregatício ou Funcional	19758	13046	6712	23755	12612	11143	4431

⁵⁸ Código Fonte: Impacto do Financiamento pelo cálculo simples de Diferença em Diferenças.sql

⁵⁹ Tabela completa: Impacto por meio da Diferença em Diferenças simples.xlsx

Apresentação de Trabalho Técnico	20253	10959	9294	27886	14897	12989	3695
Início de Projeto de Pesquisa	5786	3198	2588	11184	5991	5193	2605
Participação em Congresso	10990	7136	3854	14957	8677	6280	2426
Fim de Vínculo Empregatício ou Funcional	14350	9544	4806	17351	10190	7161	2355
Início de Participação em Projeto	4923	2877	2046	9583	5532	4051	2005
Fim de Projeto de Pesquisa	3975	2143	1832	7740	4048	3692	1860
Produção de Trabalho Técnico	3068	1869	1199	4506	1635	2871	1672
(...)							
Música	1728	1240	488	430	248	182	-306
Outras Orientações Concluídas	15436	6913	8523	10760	3178	7582	-941
Participação em Banca de Graduação	15346	6480	8866	10048	3041	7007	-1859

Fonte: Autoria própria, com base no Lattes e no portal de dados abertos do CNPq (2023).

Exemplificando, a soma de todos os artigos publicados para quem não recebeu bolsas e até o ano de 2013 foi de 2.991, enquanto a soma em 2019 foi de 10.446. Portanto, a diferença entre os anos de 2014 e 2019 indica esse grupo publicou $10.446 - 2.991 = 7.455$ artigos. Já o grupo que recebeu bolsas do CNPq publicou, calculando as diferenças de publicações nesse mesmo período, $18.442 - 4.165 = 14.277$ artigos. Assim, a diferença dessas diferenças é $14.277 - 7.455 = 6.822$, o que indica que a concessão de bolsas a 2.296 pesquisadores teve um impacto de uma média de 3 artigos a mais por pesquisador. Essa metodologia é conhecida como “Diferença em Diferenças”.

Esse método de cálculo é simples e útil, mas não tem o poder estatístico de uma regressão linear. Assim, como o CNPq possui informações sobre o montante financiado em bolsas de doutorado para cada doutorando, adotou-se a metodologia de econometria de regressão linear, usando o modelo de diferença em diferenças e que buscou calcular o efeito desse financiamento, de modo que a Equação 12 foi alterada para:

Equação 19 – Medida do Impacto do CNPq por meio de Regressão multivariada

$$Indicador_{2019} = \alpha + \beta Indicador_{2013} + \theta Financiamento_{CNPq}$$

Na equação acima, $Indicador_{2013}$, $Indicador_{2019}$, α e β mantêm o significado original (soma dos indicadores até os respectivos anos, constante e influência de um indicador no outro com o tempo). $Financiamento_{CNPq}$ dá a soma total investida pelo CNPq ao bolsista em bolsas de doutorado, tanto no país quanto no exterior, entre os anos de 2014 e 2019. Assim, θ dá a medida de impacto desse financiamento, de modo que $\theta Financiamento_{CNPq}$ dá o efeito desse em $Indicador_{2019}$, sendo, portanto, medida de efetividade.

Realizando as diversas regressões acima⁶⁰, filtrando-se os indicadores não significativas (Prob (F-statistic) < 0,01; $P > |t| < 0,01$, amostra > 200, $R^2_{ajust} > 0.3$), podemos chegar na seguinte tabela:

Tabela 11 - Impacto de Bolsas do CNPq aos Doutorandos 201461.

Indicador	Nº	R ² ajustado	α	β	Impacto	Erro Padrão (+/-)
Artigo Publicado	4148	45%	3,43	1,93	1924	339
Início de Vínculo Empregatício ou Funcional	4695	61%	1,98	1,29	1922	241
Fim de Vínculo Empregatício ou Funcional	4494	70%	1,79	1,19	643	165
Participação em Congresso	3996	69%	1,75	1,20	586	194
Participação em Encontro	3334	77%	0,71	1,17	367	140
Início de Residência Médica	142	99%		1,00	239	66
Orientação em Andamento de Graduação	436	34%	1,13	1,15	-1034	386
Participação em Banca de Graduação	2561	54%	6,39	1,18	-3274	950
Outras Orientações Concluídas	2414	50%	7,22	1,10	-3438	1018

Fonte: A autoria própria com base no Portal de Dados Abertos do CNPq (2022).

Observa-se que a Tabela 13 apresenta impacto medido muito menor que o da Tabela 10. Pode-se considerar que a Tabela 10 dá medida de impacto sem descarte estatístico, enquanto a Tabela 13 dá apenas o impacto com alta probabilidade estatística. Já a Tabela 9 dá o Impacto sem levar em conta o valor financiado. Dessa forma, o impacto real deve estar entre essas tabelas, considerando-se o da Tabela 10 sobrestimado e o da Tabela 13 subestimado, mas que, provavelmente, o modelo mais próximo da realidade seja o da Tabela 10.

14.3.8. Análise Custo-Efetividade

A Análise custo-efetividade, que é muito aplicada em economia e bastante utilizada para permitir a comparação entre distintas políticas públicas, busca comparar aquilo que se obteve com o custo de se obter esse resultado. Considerando que o custo é o total investido pelo CNPq em bolsas de doutorado, é constante no valor de R\$ 289.924.589,50, de modo que podemos dividir esse valor pelo impacto medido para ter a medida de custo-efetividade. Análise

⁶⁰ Código Fonte: 17.1 - Regressões Lineares - CNPq e Pagtos, Sem preencher nem best_r2.ipynb

⁶¹ Resultado completo em Regressões variáveis_independentes ['qty_2014'^J 'percent_pagos'] confianças [0.05] 2014 Pareamento False best_r2False CNPq True CAPES False Ambos False Nenhum True preenche_todos_idsFalse.xlsx.

semelhante não pôde ser realizada com a CAPES pois essa não forneceu informações financeiras.

Utilizando a mesma metodologia anterior ($\text{Prob}(F\text{-statistic}) < 0,01; P > |t| < 0,01$, amostra > 200 , $R^2_{\text{ajust}} > 0,3$) e unindo indicadores com “Início” e “Fim” para que seja construído o indicador “Novo”:

Tabela 12 - Custo-efetividade de bolsas de doutorado pelo CNPq a quem iniciou seu doutorado em 2014.

Indicador	Impacto	Custo-Efetividade
Participação em Banca de Graduação	-2768	-R\$ 96.240,32
Outras Orientações Concluídas	-2945	-R\$ 90.475,31
Artigo Publicado	2053	R\$ 129.746,04
Tradução	1978	R\$ 134.663,90
Novo Vínculo Empregatício ou Funcional	1294	R\$ 205.887,89
Participação em Exposição	664	R\$ 401.391,35
Participação em Banca de Mestrado	-643	-R\$ 414.532,95
Conclusão de Aperfeiçoamento	441	R\$ 604.521,22
Nova Atividade de Ensino	363	R\$ 734.066,58
Participação em Oficina	265	R\$ 1.005.679,58
Início de Residência Médica	200	R\$ 1.331.548,51

Fonte: Autoria própria (2023).

Finalmente, podemos comparar os resultados anteriores (em que foi verificando se houve ou não bolsas pela CAPES ou pelo CNPq - Tabela 9), aos resultados da Tabela 12, chegando-se na seguinte tabela:

Tabela 13 – Custo-Efetividade da concessão e Bolsas a Doutorandos 2014 por concessão ou não e por valor investido.

Indicador	Custo-Efetividade por Concessão ou não	Custo-Efetividade por Pareamento	Média
Participação em Banca de Graduação	-R\$ 101.514,21	-R\$ 88.553,63	-R\$ 95.033,92
Outras Orientações Concluídas	-R\$ 122.227,90	-R\$ 84.329,43	-R\$ 103.278,67
Artigo Publicado	R\$ 69.643,19	R\$ 150.688,46	R\$ 110.165,82
Novo Vínculo Empregatício ou Funcional	R\$ 129.027,41	R\$ 226.680,68	R\$ 177.854,04
Orientação em Andamento de Aperfeiçoamento ou Especialização	#N/D	-R\$ 133.667,40	-R\$ 133.667,40
Trabalho em Eventos	R\$ 157.225,92	#N/D	R\$ 157.225,92
Detalhamento do Trabalho	R\$ 172.266,54	#N/D	R\$ 172.266,54

Orientação em Andamento de Graduação	#N/D	-R\$ 280.391,29	-R\$ 280.391,29
Conclusão de Curso de Curta Duração	R\$ 432.078,37	#N/D	R\$ 432.078,37
Início de Curso de Curta Duração	R\$ 435.322,21	#N/D	R\$ 435.322,21
Ministério de Curso de Curta Duração	R\$ 502.468,96	#N/D	R\$ 502.468,96
Participação em Congresso	R\$ 545.997,34	R\$ 494.751,86	R\$ 520.374,60
Novo Projeto de Pesquisa	R\$ 677.393,90	#N/D	R\$ 677.393,90
Participação em Encontro	#N/D	R\$ 789.985,26	R\$ 789.985,26
Nova Atividade de Ensino	R\$ 821.316,12	#N/D	R\$ 821.316,12
Publicação de Capítulo de Livro	-R\$ 929.245,48	#N/D	-R\$ 929.245,48
Início de Residência Médica	#N/D	R\$ 1.213.073,60	R\$ 1.213.073,60

Fonte: Autoria própria (2023).

Conforme já explicado, a regressão por pareamento busca eliminar possível viés de seleção e, por isso, é mais precisa. Dessa forma, como medida de custo-efetividade, deve-se considerar, preferencialmente, essa forma de regressão em relação à da Tabela 9. De qualquer forma, observa-se que as medidas, quando disponíveis em ambas as formas, não se afastam muito uma da outra, o que comprova a robustez do método aplicado.

14.4. Conclusão

O estudo confirmou a importância da concessão de bolsas para a formação no nível de doutorado. Ainda, a análise cientométrica realizada deu medida estatística do impacto da concessão de bolsas, tanto pela CAPES, quanto pelo CNPq, de três distintas formas.

Na Tabela 13, os dois primeiros impactos são negativos. Mas, ambos são impactos que se traduzem em atividades de pouca especialização: Participação em Banca de Graduação e Outras Orientações Concluídas. Em seguida, vemos Artigos Publicados, com a custo-efetividade de R\$ 110.165,82 por artigo. O custo por artigo não é alto se pensarmos que um artigo significa um avanço na fronteira do conhecimento e a introdução de mais um pesquisador brasileiro. Ainda, a bolsa de doutorado do CNPq possui o valor de R\$ 148.800,00 (R\$ 3.100,00 mensais à época), o que se traduz num impacto não só de um artigo por doutor, mas de um novo vínculo e um trabalho publicado em evento por doutorando.

O estudo é capaz de indicar impactos negativos, como a publicação de capítulos de livros, de modo a indicar pontos que devem ser incentivados aos doutorandos. Por exemplo, pode-se

incentivar que doutorandos orientem graduandos ou que também publiquem livros, ao invés de só artigos, de modo a reduzir esse impacto negativo e possibilitar uma formação mais ampla por parte do doutorando.

O estudo permite a comparação entre diversas políticas públicas dando indicação de qual política possui uma efetividade melhor e, portanto, servindo como orientação. Alerta-se, no entanto, que sem uma análise mais detalhada a metodologia não consegue indicar qual a “melhor” política pública. Aliás, essa é justamente a grande vantagem, mas também a grande limitação do estudo: por ser quantitativo, é escalável a diversas políticas públicas, facilmente aplicado uma vez desenvolvido e comparável no tempo, mas também incapaz de dar um *insight* mais aprofundado e indicar os mecanismos por trás desse impacto (ou da falta dele), bem como as consequências da política pública de forma mais ampla.

Por isso, sugere-se continuidade da pesquisa a outras políticas públicas e também na forma de séries temporais, bem como estudos de caso que poderão confirmar o impacto e indicar as mudanças sociais decorrentes desse. Além disso, seria interessante confirmar quais das medidas de impacto é a mais precisa.

Um estudo de caso em potencial que já se sugere é a correlação entre doutorado e novos vínculos. O impacto negativo no grupo de controle sugere a desistência do vínculo atual. Por outro lado, o forte impacto positivo em geral sugere que as bolsas de estudo são um excelente remédio ao problema.

Em consideração final, os investimentos praticados, tanto pela CAPES quanto pelo CNPq, no geral, têm impacto positivo e pode-se dizer que proporcionaram, aproximadamente, a publicação de 15.019 artigos, 8.927 trabalhos em eventos, 7.198 novos vínculos, 2.060 trabalhos técnicos, 1.807 novas atividades de ensino, entre diversos outros, o que colabora com as metas do PNE e com o desenvolvimento socioeconômico do País.

15. ANEXO III – ARTIGO: O IMPACTO DE UM INVESTIMENTO PÚBLICO EM CIÊNCIA NA FORMAÇÃO DE RECURSOS HUMANOS EM NÍVEL DE MESTRADO E DOUTORADO: O CASO DO UNIVERSAL 2014

An impact of an public investment in science into human resources formation at masters and doctoral levels: The case of Universal 2014

Abstract: In this article we present an assessment of an investment in scientific research (Universal 2014) carried out by the Brazilian government through the National Council for Scientific and Technological Development (CNPq). Its impact on education was estimated by measuring formation of masters and PhDs. We used multivariate regression in a quasi-experiment, measuring the correlation between the amount invested in Universal 2014 and the number of formations. The methodology was recommended by the World Bank and the data taken from the Lattes Curriculum, a public curriculum maintained by the Brazilian government. There was a high impact on human resource formation, indicating the need for greater investment by the government and showing a relationship between investment in science and improvement in education. Finally, a list of the possible impact of other indicators is presented, pointing to an impact far above the amount invested.

Keywords: Impact Assessment. Human Resources Training. Quasi-Experiment. Curriculum Lattes. Public Policy for Investment in Science.

O Impacto de um investimento público em ciência na formação de recursos humanos em nível de mestrado e doutorado: o caso do Universal 2014

Resumo: Neste artigo apresentamos uma avaliação de um investimento em pesquisa científica (Universal 2014) realizado pelo governo brasileiro por meio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Seu impacto na educação foi estimado pela mensuração da formação de mestres e doutores. Utilizamos a regressão multivariada em um quase-experimento, medindo a correlação entre o valor investido no Universal 2014 e o número de orientações concluídas. A metodologia foi recomendada pelo Banco Mundial e os

dados retirados do Currículo Lattes, currículo público mantido pelo governo brasileiro. Houve alto impacto na formação de recursos humanos, indicando a necessidade de maior investimento por parte do governo e mostrando uma relação entre investimento em ciência e a melhoria na educação. Por fim, é apresentada uma lista do possível impacto de outros indicadores, apontando para um impacto muito acima do valor investido.

Palavras-chave: Avaliação de Impacto. Formação de Recursos Humanos. Quase-Experimento. Currículo Lattes. Política Pública de Investimento em Ciência.

Un impacto de la inversión pública en ciencia en la formación de recursos humanos a nivel de maestría y doctorado: el caso de Universal 2014

Resumen: En este artículo presentamos una evaluación de una inversión en investigación científica (Universal 2014) realizada por el gobierno brasileño a través del Consejo Nacional de Desarrollo Científico y Tecnológico (CNPq). Su impacto en la educación se estimó midiendo la formación de maestrías y doctorados. Utilizamos regresión multivariante en un cuasiexperimento, midiendo la correlación entre el monto invertido en Universal 2014 y el número de formaciones. La metodología fue recomendada por el Banco Mundial y los datos tomados del Currículo Lattes, un currículo público mantenido por el gobierno brasileño. Hubo un alto impacto en la formación de recursos humanos, lo que indica la necesidad de una mayor inversión por parte del gobierno y muestra la relación entre la inversión en ciencia y la mejora en la educación. Finalmente, se presenta una lista del posible impacto de otros indicadores, apuntando a un impacto muy por encima del monto invertido.

Palabras clave: Evaluación de Impacto. Formación de Recursos Humanos. Cuasiexperimento. Curriculum Lattes. Política Pública de Inversión en Ciencia.

15.1. Introdução

O Plano Nacional de Educação (PNE), [Lei nº 13.005/2014](#), traz em suas metas 14 e 16 “Elevar gradualmente o número de matrículas na pós-graduação de modo a atingir a titulação anual de 60.000 (sessenta mil) mestres e 25.000 (vinte e cinco mil) doutores” e “Formar, em nível de pós-graduação, 50% (cinquenta por cento) dos professores da educação básica, até o último ano de vigência deste PNE, e garantir a todos (as) os (as) profissionais da educação básica formação continuada em sua área de atuação, considerando as necessidades, demandas

e contextualizações dos sistemas de ensino”. Apesar das metas indicadas, Cabrera e Araújo (2018) verificaram uma diminuição no número de mestrados profissionalizantes, tanto em matrículas quanto em formações, levantando a necessidade de políticas públicas que respondam aos desafios atuais e atendam as crescentes demandas da educação nacional. Essa diminuição também pode estar ocorrendo no mestrado acadêmico e no doutorado.

Uma das formas de atender as metas do PNE é o fomento à formação de recursos humanos, que, de acordo com (WILBERTZ, 2013b), pode se dar direta ou indiretamente, pelo setor público ou privado. A formação indireta se dá por meio de investimento em projetos de CT&I, enquanto a direta é o financiamento a universidades e seus programas de PPG. No Brasil, o fomento público indireto é realizado por órgãos tanto estaduais quanto federais. No âmbito federal, o fomento indireto à formação de mestres e doutores é feito por meio do Ministério da Ciência, Tecnologia e Inovação (MCTI), através do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), e por meio do Ministério da Educação (ME), através da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Enquanto cabe ao CNPq “promover e fomentar o desenvolvimento e a manutenção da pesquisa científica e tecnológica e a formação de recursos humanos qualificados para a pesquisa, em todas as áreas do conhecimento” (“Competências — Português (Brasil)”, [s.d.]), cabe à Capes papel na expansão e consolidação da pós-graduação *stricto sensu* (mestrado e doutorado) em todos os estados da Federação (CAPES, 2012). Assim, enquanto compete ao CNPq, que é ligado ao MCTI, o financiamento à pesquisa, à tecnologia e à inovação, compete à Capes, ligada ao ME, o financiamento para a formação de recursos humanos. No entanto, no que diz respeito a formação de mestres e doutores, essas tarefas se confundem, uma vez que boa parte da pesquisa brasileira se liga a programas de pós-graduação.

O CNPq foi criado em 1951 e tem como missão fomentar e atuar nas políticas públicas de CT&I. Uma forma de compreender o papel do CNPq é analisar as ações de financiamento realizadas. Em 2021 o CNPq executou pagamentos no valor de R\$ 1.131.524.331,05 (um bilhão de reais)⁶². Até o ano de 2016, o CNPq havia lançado 975 ações do tipo “Edital” e 549 do tipo “Bolsas e Auxílios”⁶³. Dessas, três ações se destacam, tanto pelo volume total solicitado na demanda quanto no volume de recursos financiados. Os Editais Universais representaram 14,36%⁶⁴ do total (R\$ 17.012.482.313,00), os Institutos Nacionais de Ciência e Tecnologia

⁶² <https://www.portaltransparencia.gov.br/orgaos/20501-conselho-nacional-de-desenvolvimento-cientifico-e-tecnologico>

⁶³ Relatório tirado por meio de uma ferramenta interna ao CNPq que foi descontinuada em 2016.

⁶⁴ Os valores apresentados são a demanda bruta, ou seja, o total de recursos solicitados. O total aprovado é muito inferior, mas segue aproximadamente a mesma proporção.

obtiveram 11,65% (R\$ 13.793.092.584,00) e Bolsas Produtividade ficaram com 10,04% (R\$ 11.890.687.902,00). Essas três ações também representam três formas de atuação do CNPq: investimento em projetos de CT&I, investimento direto em grandes centros de tecnologia e concessão de bolsas.

Ocorre que a formação de recursos humanos exigido pelo PNE pode resultar de todas as três formas de investimento acima citados. Políticas públicas trazem efeitos tanto diretos quanto indiretos. Os efeitos indiretos estão de acordo com a teoria da Hélice Tripla em que, de acordo com Leydesdirff e Ivanova (2016), há três eixos, ortogonais, (1) relativo à produção do conhecimento, (2) relativo à geração de riquezas (comércio e indústria) e (3) relativo ao controle normativo (governância), de modo que cada vértice formado pode tanto inibir quanto catalisar alterações nos outros dois vértices, em sistema dinâmico e autogerido. Não é somente a educação que afeta a ciência e o desenvolvimento, mas a ciência também afeta a educação, tanto diretamente, quanto indiretamente, sendo que a governância ainda é afetada e afeta os outros eixos. Assim, tanto o investimento em projetos de CT&I, quanto o investimento direto em grandes centros de tecnologia, podem trazer resultados indiretos em termos de formação de recursos humanos, tanto mestres quanto doutores.

O efeito de investimentos em projetos de CT&I na formação humana é conhecido e esperado pelo CNPq. No questionário de Prestação de Contas Final⁶⁵ das chamadas Universal, um dos itens de preenchimento obrigatório é justamente a formação realizada, em termos de iniciação científica, mestres, doutores. O CNPq espera o mesmo resultado indireto com as outras duas ações anteriormente citadas: investimentos a grandes centros de tecnologia e concessão de bolsas de Produtividade em Pesquisa. Assim, o CNPq espera como parte dos resultados da política pública de investimento em CT&I a formação de mestres e doutores.

No Brasil, a avaliação das políticas públicas, inclusive com divulgação do objeto a ser avaliado e dos resultados alcançados, é obrigação imposta pela Constituição Brasileira (art. 37, § 3º-I, § 16, art. 74, I e II, art. 165, § 16, art. 193). Na educação, a Lei de Diretrizes e Bases da Educação Nacional (LDB) assegura a avaliação nacional da educação em todos os seus níveis. No investimento em CT&I, o Novo Marco Legal (lei nº 13.243/2016) prevê prestação de contas simplificada por parte dos pesquisadores e o decreto nº 9.283, que a regulamenta, institui avaliações de resultados como um de seus princípios básicos.

A legislação brasileira não só exige que avaliações sejam realizadas para todas as

⁶⁵ Os questionários não estão publicamente disponíveis, mas podem ser adquiridos por meio da Lei de Acesso à Informação.

políticas públicas de investimento, mas também exige que essas sejam realizadas com transparência e objetividade. A transparência é princípio da Constituição Federal e também regulamentado pela Lei 12.527, a Lei da Transparência. Nela, é garantido o direito de acesso à informação de forma objetiva, transparente, clara e em linguagem de fácil compreensão.

Desta forma, esta pesquisa busca uma avaliação do efeito indireto do investimento em CT&I, particularmente da Chamada Universal 2014 do CNPq, na formação de mestres e doutores. Mais especificamente, busca-se medir a correlação entre o valor investido em um projeto de pesquisa da Chamada Universal 2014 e a formação humana resultante, ou seja, uma forma de impacto desse valor investido. Secundariamente, busca-se comparar a correlação medida com os investimentos diretos realizados na forma de bolsas, de modo a avançar no tema sobre práticas de governança na avaliação do investimento realizado na pesquisa científica, atendendo a necessidade legal de avaliação com transparência. Finalmente, pretende-se indicar se há outras correlações, indicando outros possíveis impactos e futuras pesquisas no tema a realizar.

Na introdução deste artigo apresentou-se a contextualização do tema estudado, a motivação do estudo e os objetivos da pesquisa. A seção 2 apresenta a metodologia. A seção 3 apresenta o Currículo Lattes, de onde a maior parte dos dados foram retirados, e a 4 detalha a forma de obtenção dos indicadores necessários e os procedimentos adotados para a extração dos dados. A seção 5 dá os resultados da pesquisa e, por fim, na 6 temos as considerações finais e sugestões para trabalhos futuros.

15.2. Metodologia

Segundo (BORNMANN, 2013a), o impacto da ciência é a avaliação do retorno social, cultural, ambiental e econômico, de forma que o retorno social representa a contribuição ao capital social da nação (estimulando, por exemplo, novas abordagens a problemas sociais, debates públicos ou melhorando a política pública), o retorno cultural é a contribuição para a preservação ou ganho do patrimônio cultural, como, por exemplo, o entendimento de como as culturas ou sociedade se relacionam, o retorno ambiental adiciona ao capital natural da nação (redução de lixo ou poluição, reciclagem, redução do desmatamento ou produção de energia limpa) e o retorno econômico estaria relacionado a, por exemplo, capital humano (habilidades, técnicas, produtividade e conhecimento). Dessa forma, eventual correlação entre uma política pública de investimento em CT&I e a formação de recursos humanos pode ser considerado como um dos impactos dessa política.

Para medir esse impacto, seguiu-se a metodologia indicada pelo Banco Mundial, em seu manual intitulado Avaliação de Impacto na Prática (GERTLER et al., 2018). Nele, o impacto de uma política pública é medido pela diferença entre o estado atual daquele que recebeu o tratamento (neste caso, a política pública de investimento em ciência, ou seja, o financiamento) e aquele estado de como estaria esse mesmo indivíduo caso essa política nunca tivesse existido (o contrafactual).

No entanto, o contrafactual é apenas teórico, uma vez que é impossível afirmar qual seria o estado do indivíduo sem a política e, por isso, precisa ser estimado (GERTLER et al., 2018). A forma mais básica de se estimar o contrafactual é supor que, sem o tratamento, não haveria alterações a serem medidas. Ou seja, supõe-se que o estado se manteria no tempo e, para calcular o impacto, basta verificar o estado atual daquele que recebeu o tratamento e comparar com o seu estado anterior. Assim, mede-se a mudança que o tratamento causou no tratado, supondo que o contrafactual seria estado idêntico ao do beneficiado antes de eventual financiamento. Ocorre que, ao longo do tempo, outros fatores externos ao tratamento podem afetar a pessoa tratada, fazendo com que a diferença entre os estados atual e inicial não seja devido apenas ao efeito do tratamento. Assim, o contrafactual apresenta viés, que é a diferença entre o estado estimado e o estado que seria o real do tratado (GERTLER et al., 2018).

Dessa forma, para aumentar a precisão da medida do impacto, utiliza-se um grupo de controle, um grupo que não recebeu o financiamento e que permite que se estime outros fatores externos. A escolha de quem pertencerá a esse grupo, no entanto, é complexa e pode definir o estudo realizado. De acordo com o Banco Mundial, a melhor metodologia é a experimental, em que escolher-se-ia todos os elegíveis para o tratamento e aplicar-se-ia o mesmo apenas em alguns indivíduos aleatoriamente selecionados. A seleção aleatória tem a função de garantir que, na média, estima-se corretamente todas as mudanças externas ao tratamento realizado. Subtraindo esses efeitos ao grupo de tratamento, temos o impacto da política pública. Essa é a metodologia experimental.

No entanto, como se trata de análise ex-post, não é possível a seleção aleatória e o grupo de controle precisa ser substituído por uma estimativa estatística. Uma opção seria calcular a média dos indicadores dos grupos de tratamento e controle, no período inicial e final indicados, para depois apenas se subtrair a média desses indicadores. O método tem cálculo simples e dá o impacto médio de forma direta, mas tem a desvantagem de não indicar o grau de confiança estatística, ou seja, possui uma maior chance de nos dar indicador de impacto superestimado ou irreal.

É por esse motivo que se adotou a metodologia indicada pelo Banco Mundial (GERTLER et al., 2018) da regressão multivariada. Nessa metodologia, utiliza-se de ferramentas estatísticas para se criar um grupo de controle artificial. O método supõe que todos os fatores externos à política pública que afetaram o grupo de tratamento também afetaram o grupo de controle, e vice-versa. É, portanto, um método quase-experimental.

Na regressão multivariada, uma variável busca medir a diferença, no tempo, entre o grupo que recebeu e o que não recebeu o financiamento. Uma vez que essa diferença é tratada por uma variável independente, a segunda variável passa a medir o efeito do tratamento.

A regressão linear multivariada possui diversos benefícios sobre a diferença simples, por usar de métodos estatísticos que dão a indicação da real representatividade do indicador, eliminando resultados que poderiam não ser reais. Dessa forma, este estudo seguirá o modelo adotado por Milesi et al. (2014), que usaram de regressão linear multivariada para comparar a produtividade de pesquisadores, antes e depois de um financiamento (REESE Award). Milesi et al. (2014) deixam claro que, quando possível, a melhor metodologia compara quem concorreu à mesma política pública mas não foi contemplado (grupo de controle) com quem concorreu e foi selecionado (grupo de tratamento). O Banco Mundial propõe metodologia semelhante, assim, este estudo adota essa comparação.

Milesi et al. (2014) se utilizam da seguinte equação:

Equação 20 – Modelo de Regressão Linear para comparar a Produtividade no Tempo por meio de um Estudo Bibliométrico (MILESI et al., 2014).

$$y_{ti} = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 after + \beta_4 after^2 + \sum_c \beta_c x_{ci} + u_i + r_{ti}$$

Na Equação 1, i se refere a um índice de beneficiados, t é o tempo, $after$ é o tempo após a política pública, $\sum_c \beta_c x_{ci}$ são covariáveis incluídas no modelo, u_i é variação aleatória entre beneficiados que não é uma função do tempo e r_{ti} é variação aleatória ano a ano dos pesquisadores financiados ($u_i + r_{ti}$ estimam o erro da equação).

Ocorre que Milesi et al. (2014) se utilizam de uma série temporal, enquanto este estudo compara apenas dois pontos no tempo. A análise temporal não é necessária, uma vez que são usados apenas dois pontos temporais de referência necessários para a análise do impacto (o inicial e o final). Dessa forma, as variáveis independentes quadráticas e as constantes que refletem seus erros, ou seja, as variáveis que buscam estimativa de progressão temporal e que refletem as alterações temporais não lineares, não serão aqui necessárias.

Além disso, Milesi et al. (2014) adotam soma de variáveis correlatas que ajustam para

ano de doutorado, financiamento por outras agências de fomento, nota da instituição de vínculo, área de estudo e o sexo. Muitos desses dados não estão disponíveis e, portanto, não foram diretamente utilizados. No entanto, entende-se que todas essas variáveis estão altamente correlacionadas com o nível do indicador antes do financiamento. Dessa forma, a equação será alterada para:

Equação 21 – Modelo de Regressão Linear sem variáveis temporais.

$$y_{ti} = \beta_0 + \beta_3 after + \beta_c x_{ci} + u_i$$

Finalmente, β_0 e u_i são constantes, que representam não só o valor inicial do estado, mas também o eventual erro. Dessa forma, as duas podem ser unidas em apenas uma, uma vez que buscamos apenas o β_3 (o impacto) na equação acima. Simplificando a equação, temos:

Equação 22 – Modelo de Regressão Linear simplificada.

$$y_{ti} = \beta_i + \beta_1 before_i + \beta_3 after_i$$

Em que $\beta_i + \beta_1 before_i$ mede a correlação de fatores externos ao financiamento e substitui $\beta_0 + \sum_c \beta_c x_{ci} + u_i + r_{ti}$.

Rearranjando a equação é alterando os nomes das variáveis, no intuito de facilitar a compreensão da fórmula utilizada, a regressão linear multivariada tem a seguinte equação:

$$Indicador_{atual} = C + I * Indicador_{anterior} + P * Financiamento$$

Equação 23 - Regressão Linear entre índice atual, índice anterior e financiamento.

Na equação acima, *Indicador* representa a variável estudada, como por exemplo, orientação de doutorado. *C* é uma constante que reflete o estado inicial e também inclui o erro, *I* dá a correlação de fatores externos não temporais, entre o *Indicador_{anterior}* e o *Indicador_{atual}* e *P* mede a correlação do financiamento com o *Indicador_{atual}*.

Como *Indicador* é, neste estudo, o número de mestres e doutores orientados, tomando o número de doutores orientados como indicador de exemplo, *Doutores₂₀₁₄* é o número de doutores formados pelo pesquisador até o final do ano de 2014 e *Doutores₂₀₁₉* esse número até o final de 2019.

C representa uma constante e estima o contrafactual quando se supõe que não houve fatores externos ou temporais. Dessa forma, estimar *C* representaria estimar a média de doutores formados no ano de 2014, ou seja, o estado inicial do indicador, mais o erro.

I representa a taxa constante de progressão no tempo independentemente do financiamento. Ou seja, representa a tendência que se tem de crescimento espontâneo do

número de doutores formados pelos pesquisadores, ou a medida de efeitos externos. Assim, $I * Indicador_{anterior}$ estima os efeitos externos e temporais no contrafactual.

P mede o efeito do financiamento, ou seja, a correlação entre o valor investido e o número de doutores formados. Como o impacto é a medida da diferença entre os estados dos grupos de tratamento e controle, $P * Financiamento$ nos dá justamente a estimativa do impacto da política pública no indicador em questão.

15.3. Currículo Lattes

Para a realização deste estudo, precisou-se extrair informações sobre a formação humana realizada de cada pesquisador. Esses foram retirados dos Currículos Lattes, que é um currículo informatizado, padronizado, público e mantido pelo governo brasileiro, mais especificamente pelo CNPq. Ele é de uso obrigatório, no Brasil, para todos aqueles que queiram concorrer a financiamentos públicos para uma pesquisa científica e se tornou um padrão no registro da vida dos estudantes de pós-graduação e pesquisadores do país. Além disso, ele é hoje adotado pela maioria das instituições de fomento, universidades e institutos de pesquisa do País, sendo adotado, inclusive, por instituições particulares que o usam como padrão para vagas de emprego, financiamentos ou outras vantagens. Ele é de acesso público, tanto para o cadastro quanto para a consulta e permite a consulta por área, formação, atuação profissional, palavras chaves, entre outros.

Ainda, por ter uma estrutura padronizada, permite que se realize mineração de dados para que os indicadores sejam automaticamente recuperados. Para que se tenha uma compreensão da dimensão de informações presentes no Currículo Lattes, o arquivo que representa sua estrutura possui 6.574 linhas, ou seja, são mais de 5.000 diferentes campos, padronizados, que podem ser preenchidos e recuperados. Observa-se que, no momento da escrita deste artigo, a Plataforma Lattes tem 7.427.277 currículos cadastrados, que após importados geram cerca de 1 Tb de dados.

Como vamos medir a mudança ao longo do tempo do grupo de controle e do grupo de tratamento, procuramos apenas aqueles campos que podem ser diretamente relacionados com o ano. Sob essas regras, extrai-se do Currículo Lattes 140 indicadores diferentes. Para que se entenda melhor, damos o exemplo de um doutorado orientado. São mais de 30 diferentes informações que podem ser inseridas e catalogadas, como nome do doutorando, instituição, título da tese, coorientadores etc. No entanto, cada doutorado orientado possui apenas um ano

de titulação. Portanto, de todas as informações que se pode extrair de uma entrada dessas no Currículo Lattes, tira-se apenas um indicador, que foi nomeado de “Orientação de Doutorado”.

Ressalta-se que, sob estas regras, pode-se extrair 140 indicadores do Currículo Lattes. No entanto, apenas quatro indicadores foram utilizados inicialmente: Doutorados Orientados, Mestrados Orientados, Mestrados Profissionalizantes Orientados e Pós-Doutorados Orientados.

Pelo volume de informações que devem ser analisadas, este estudo se baseia no uso de BigData e de acordo com Guthrie et al (2013) pertencem ao que ele chama de ferramentas de análise do Grupo 2: bibliometria, mineração de dados e análise econométrica. O Grupo 2 constitui de Ferramentas Quantitativas e são normalmente escaláveis, transparentes, permitem comparações entre avaliações diferentes e são adequadas a usos frequentes e longitudinais. São normalmente livres de julgamento, mas a implementação da avaliação costuma ter alto custo inicial. Assim, de acordo com Guthrie et al (2013), o estudo pode ser facilmente expandido a outros indicadores ou outros financiamentos. Assim, como parte dos objetivos secundários, o estudo foi expandido a outros indicadores, revelando diversos indicadores que possuem correlação com o investimento realizado. Essa análise é apresentada nas considerações finais.

15.4. Obtenção e tratamento dos dados

Antes de se extrair quaisquer dados, foi necessário definir os anos de início e de fim a serem considerados. Stern (2014) demonstra que a correlação entre o quantitativo de citações ao longo do tempo já possui forte correlação ao número de citações em anos anteriores a partir do terceiro ano de publicação (janela de dois anos após o ano de publicação do artigo), explicando cerca de 70% (69,2% para economia e 71,8% para ciência política) dos dados (STERN, 2014). A correlação aumenta significativamente nos anos posteriores, para cerca de 85% com uma janela de 3 anos, e cerca de 95% em janela de 4 anos. Como o Universal tem uma duração de dois anos, foi adotada a janela de seis anos para a medida do impacto. Ressalta-se, no entanto, que essa janela foi escolhida com base na correlação de um único indicador e que um maior estudo seria necessário para definir se essa está adequada.

No entanto, no ano de 2020 houve fato singular que pode representar um viés capaz de prejudicar os resultados encontrados, a pandemia COVID-19, muito embora a metodologia busque justamente eliminar o efeito de fatores externos como esse. Mesmo assim, para evitar que a pandemia interfira na aplicação do método, foi adotado como o tempo final do estudo o dia 01/01/2020, de forma que todos os Currículos Lattes que não foram atualizados dessa data

a janeiro de 2022, data de aquisição dos currículos, foram desconsiderados.

Com uma janela de seis anos, o tempo inicial é o ano de 2014, assim, o Edital Universal 2014 foi escolhido como objeto de estudo. Observa-se que a escolha do ano final de 2019 (ou seja, todos os indicadores anteriores a 01/01/2020) como objeto de estudo também se deu para que se tenha um maior número de Currículos Lattes atualizados, uma vez que apenas esses podem ser usados. Assim, quanto mais recente é o estudo, menos currículos estão atualizados e menor é o poder estatístico.

Uma vez definido a chamada a ser estudada e os anos a serem considerados, passou-se à coleta de dados. Foram coletadas três origens de dados diferentes: O Currículo Lattes de todos os pesquisadores envolvidos, a Planilha de Demanda Bruta, que lista todos os pesquisadores que concorreram à chamada estudada e a listagem de todos os pagamentos realizados pelo órgão.

Os Currículos Lattes são publicamente acessíveis. No entanto, ferramentas de proteção impedem a extração automática por meio de web scraping (programa de computador que baixa os currículos automaticamente). Dessa forma, os currículos foram obtidos por meio da ferramenta Extrator Lattes, disponibilizado pelo CNPq (“Extração de dados - Portal Memória”, [s.d.]). Para a extração, foi necessária a criação de um programa em Python que verificou os dados necessários para a extração, realizou a extração e a coleta dos indicadores, em cada um dos currículos baixados.

A Planilha de Demanda Bruta foi retirada por meio de um sistema interno ao CNPq de relatórios. No entanto, a planilha também pode ser obtida por meio de solicitação ao canal da Lei de Acesso à Informação. Nela, constam o nome completo de cada pesquisador que concorreu ao Universal 2014, bem como diversos outros dados relativos à proposta submetida, incluindo-se o número do processo, realizado para identificar se houve ou não pagamentos em relação ao processo submetido.

A listagem de todos os pagamentos realizados está publicamente disponível em http://dadosabertos.cnpq.br/pt_BR. A planilha contempla o número do projeto, o valor pago e o ano de pagamento, além de outras informações. Dessa forma, liga-se o pagamento ao pesquisador pelo número do projeto, e o pesquisador a seu Lattes por meio do nome completo. Finalmente, para evitar a possibilidade de homônimos, foi realizada verificação por meio de outra ferramenta internamente disponível de relatórios, que liga por meio de chave única o nome do pesquisador a seu Lattes. Dessa forma, garantiu-se que não houve casos de homônimos no estudo realizado.

Para a realização da regressão e uma vez que se adotou o limite imposto pela pandemia como aquele de observação final dos dados, dos 7.427.277 currículos cadastrados na Plataforma Lattes, foram importados 2.651.713 que são aqueles atualizados a partir de 01/01/2020 (a listagem pode ser obtida no portal do Extrator Lattes). Observa-se que a importação e o tratamento desses currículos geraram mais de 700 gigabytes de dados.

Todos os 2.651.713 foram minerados em busca dos dados, para que fossem retirados os dados básicos que permitissem sua correta identificação, bem como seus respectivos indicadores. A importação foi necessária uma vez que a planilha fornecida pelo órgão não possui o número identificador do currículo, mas apenas o nome do pesquisador. Como resultado, foi construída tabela contendo 253.506.545 de linhas, cada linha indicando o ano em que foi registrado cada um dos 140 diferentes indicadores de cada currículo.

No Edital Universal 2014 houve 16.902 pesquisadores participantes, sendo que desses 15.907 tinham o currículo atualizado em 01/01/2020 e apenas 5.222 tiveram financiamento realizado. Todos os 15.907 currículos foram convertidos nos 140 diversos indicadores, calculados tanto para o ano de 2014 quanto para o final do ano de 2019, gerando, portanto, uma tabela de cinco colunas: identificador do pesquisador, indicador, pagamento, indicador em 2014 e indicador em 2019.

O resultado foi novamente trabalho em outra tabela que somou os indicadores anteriores aos anos 2014 e 2019, de modo a possibilitar a comparação no tempo, e agrupados em faixa e área de pesquisa, resultando em tabela com 3.028.200 linhas com indicadores pertencentes a 15.907 pesquisadores.

O Universal 2014 é separado em três distintas faixas de concorrência, a depender do valor solicitado: Faixa A, Faixa B e Faixa C. Dessa forma, pesquisadores que submeteram propostas para a Faixa A não concorrem com aqueles que submeteram à Faixa B, e assim por diante. Assim, os dados foram agrupados por essas faixas e também pelas 80 áreas de pesquisa em que foi permitida a concorrência, resultando em 33.600 tabelas (3 modalidades x 80 áreas x 140 indicadores).

15.5. Resultados e discussões

Uma vez levantados os dados e definidas as 33.600 tabelas, aplicou-se, então, em cada uma dessas tabelas, a regressão linear da Equação 4:

$$\text{Indicador}_{2020} = C + I * \text{Indicador}_{2014} + P * \text{Financiamento}$$

Uma das vantagens de usar a regressão multivariada é que a estatística nos dá a medida da hipótese nula “ $P > |t|$ ”, ou seja, a probabilidade estatística de que, considerados fatores aleatórios e não observados, considere-se que o indicador resultou num impacto quando isso não é verdadeiro. Em outras palavras, o teste da hipótese nula nos dá a probabilidade de que consideremos um impacto que, na verdade, não foi causado pelo financiamento e sim por outros fatores externos e/ou aleatórios. Com o teste da hipótese nula “ $P > |t|$ ”, pode-se realizar um estudo exploratório em todos os indicadores e daí se retirar quais foram aqueles com alta probabilidade estatística de representar um impacto real, além da medida desse impacto. Para facilitar a compreensão, adotaremos o conceito de Grau de Confiança, definido como $1 - “P > |t|”$ e expresso em porcentagem. Assim, o Grau de Confiança nos daria as chances da estatística ser relevante ou real.

Neste estudo, foram listados apenas os impactos com Grau de Confiança de 99% (em que o teste da hipótese nula “ $P > |t|$ ” é inferior a 0,01). Simplificadamente, tem-se assim 99% de certeza de que o impacto medido é real. Também, tem-se alta chance probabilística de que o impacto medido está subvalorizado, uma vez que muitos impactos reais não atingirão o nível de confiança proposto, isso porque, nos indicadores que não satisfizerem o grau de confiança de 99%, teremos apenas 1% de certeza de que não houve impacto nessa área, com 99% de erro. Ou seja, temos a certeza de que estamos subestimando o impacto encontrado, mas, ao mesmo tempo, estamos garantindo uma alta probabilidade estatística de que esse impacto é real. A medida foi necessária para que se reduza a possibilidade de se medir impacto não condizente com a realidade, pois o impacto é indireto e resultante de milhares de regressões.

Assim, chega-se ao impacto do programa, aqui definido e calculado como $P * \text{Financiamento}$. Ou seja, para se chegar no impacto, cada parâmetro P considerado significante (99% de Grau de Confiança) foi multiplicado pelo valor total financiado naquela faixa, programa e indicador, para se chegar ao Impacto Médio. O resultado foi novamente filtrado de modo a eliminar os impactos menores que uma unidade, uma vez que impactos muito pequenos também podem ser considerados como nulos. Finalmente, todos os impactos foram agrupados por indicador e somados, chegando-se ao impacto final da Política Pública:

Tabela 14 – Impacto da Chamada Universal 2014 em Formação Humana a nível de Pós Graduação

Indicador	Impacto
Orientação Concluída de Doutorado	604
Orientação Concluída de Mestrado	294
Orientação Concluída de Pós-Doutorado	193

Orientação em Andamento de Doutorado	91
Orientação em Andamento de Pós-Doutorado	9
Orientação em Andamento de Mestrado	-27

Fonte: Produção própria.

Chama atenção a existência de impacto negativo. No entanto, pode-se interpretar o impacto negativo no número de orientações em andamento de mestrado como, na verdade, um impacto positivo na medida em que houve migração de orientações de mestrado para o doutorado. Ou seja, enquanto o número de orientações em andamento de mestrado cai, o número de orientações em andamento de doutorado sobe significativamente, indicando que o financiamento levou os pesquisadores a darem preferência a uma formação de maior nível.

Assim, o impacto indireto do investimento realizado na educação é claro. De acordo com os dados abertos do CNPq, foi financiado nessa política pública o valor total de R\$ 194.101.318,26. Com esse investimento, notamos impacto indireto na formação de 294 mestres, 604 doutores e 193 pós-doutores. Reforça-se que esse é apenas uma pequena parcela do impacto total do financiamento, que objetivou uma pesquisa científica e não a formação de recursos humanos e que ainda teremos um aumento de 91 doutores (se todos conseguirem completar seus doutorados).

A título de comparação, o CNPq financiava diretamente a formação de recursos humanos, na forma de bolsas para Mestrado e Doutorado⁶⁶. Cursos de mestrado e doutorado são essenciais para assegurar a oferta de mão de obra qualificada, tanto para a indústria, quanto para as universidades, afetando o crescimento socioeconômico do estado (OLIVEIRA; AMARAL, 2017) e faz parte do programa educativo do Brasil.

O valor do subsídio para bolsas de mestrado é de R\$ 1.350,00, sendo a duração padrão para o financiamento de 48 meses. Assim, cada mestre formado é subsidiado em aproximadamente R\$ 32.400,00. Cada doutor recebe bolsa no valor de R\$ 2.000,00 e o subsídio tem duração padrão de 48 meses, de modo que para cada doutor formado é subsidiado o valor de R\$ 96.000,00. Já o Pós-Doutor recebia bolsa de R\$ 4.100,00 por dois anos. Assim, o valor financiado por Pós-Doutor era de R\$ 98.400,00.

No Universal 2014, temos um impacto de 267 mestres, 695 doutores e 202 pós-doutores (somando-se aqueles concluídos com os ainda em andamento). Para se obter o mesmo número de mestres, doutores e pós-doutores com bolsas, seria necessário um investimento total de R\$

⁶⁶ http://www.cnpq.br/web/guest/view/-/journal_content/56_INSTANCE_0oED/10157/25366

95.247.600,00 reais, aproximadamente 49% do valor financiado. Ou seja, o impacto indireto na educação foi de no mínimo o equivalente a 49% do valor financiado. Vale lembrar que esse valor ainda está subestimado, uma vez em que há Grau de Confiança de 99% em sua existência e que esse é apenas o impacto indireto.

A formação humana de mestres e doutores se relaciona diretamente com a qualidade da educação. Moraes et al. (2022) estudaram o Programa de Pós-Graduação em Ensino de Ciências Naturais e Matemática da Universidade Federal do Rio Grande do Norte (PPGECNM-UFRN) e indicaram o impacto do mestrado profissionalizante na educação, mostrando que materiais didáticos desenvolvidos são efetivamente usados em sala de aula em 70% dos casos, que em 55% dos casos os produtos educacionais foram publicados, que 71% propuseram atividades inovadoras, que 60% tiveram melhoria salarial, mantendo o emprego anterior em 70% dos casos. Mesmo em programas de outras áreas, espera-se efeito semelhante em que o conhecimento adquirido pelos mestres e doutores contribui para a melhoria da qualidade da educação no Brasil.

15.6. Considerações finais

A mesma metodologia foi aplicada a todos os indicadores passíveis de extração do Currículo Lattes, chegando-se a um possível impacto do Universal 2014 passível de ser medido pelo Currículo Lattes. Observa-se que, dos 140 diferentes indicadores, 26 representam o fim ou conclusão de uma atividade, com 26 respectivos indicadores que representam o início dessas atividades e que são, portanto, altamente correlacionados aos primeiros. Assim, os 11 indicadores que representam o fim de uma atividade foram agrupados com os 11 indicadores representando seu início para resultar em novas atividades, eliminando a alta correlação. Por exemplo, “Fim de Vínculos” foi subtraído de “Início de Vínculos” resultando no indicador “Novos Vínculos”.

Além disso, os 15 indicadores que representavam o início de uma atividade e que se correlacionam fortemente com seus respectivos pares que representam a conclusão dessa mesma atividade foram ignorados. Por exemplo, “Conclusão de Doutorado” é um indicador presente no estudo, mas “Início de Doutorado” foi retirado do mesmo. Restaram, portanto, 114 indicadores:

Tabela 15 – Impacto da Chamada Universal

Indicador	Impacto
Artigo Publicado	6071

Novos Vínculos Empregatícios ou Funcionais	1236
Participação em Banca de Doutorado	1119
Participação em Banca de Mestrado	899
Orientação Concluída de Doutorado	604
Trabalhos em Eventos	562
Produção de Trabalho Técnico	462
Desenvolvimento de Material Didático ou Instrucional	413
Publicação de Capítulo de Livro	321
Prêmio ou Título Recebido	316
Apresentação de Trabalho Técnico	309
Orientação Concluída de Mestrado	294
Participação em Congresso	202
Orientação Concluída de Pós-Doutorado	193
Participação em Banca de Exame de Qualificação	162
Texto em Jornal ou Revista	103
Produção de Software	96
Orientação em Andamento de Doutorado	91
Livro Publicado ou Organizado	78
Participação em Simpósio	73
Outras Participações de Bancas Julgadoras	71
Produção Bibliográfica	55
Outras Participações em Congressos	49
Patente Registrada	44
Editoração	44
Organização de Evento	38
Participação em Encontro	36
Participação em Oficina	32
Ministério de Curso de Curta Duração	32
Novas Extensões Universitárias	29
Conclusão de Extensão Universitária	26
Produção de Artes Cênicas	25
Outra produção Técnica	25
Publicação de Prefácio ou Posfácio	25
Participação em Exposição	21
Programa de Rádio ou TV	21
Participação em Banca Julgadora para Professor Titular	20
Treinamento Ministrado	14

Produção Técnica - Processos ou Técnicas	11
Produção Artística ou Cultural	11
Participação em Seminário	11
Orientação em Andamento de Pós-Doutorado	9
Tradução	8
Orientação em Andamento de Aperfeiçoamento ou Especialização	7
Novas Atividades Técnico/Científicas	6
Produção de Carta, Mapa ou Similar	4
Conclusão de Outra Formação Complementar	3
Participação em Banca Julgadora para Concurso Público	3
Produção de Artes Visuais	3
Produção - Demais Trabalhos	3
Novas Atividades de Conselho, Comissão ou Consultoria	1
Participação em Banca Julgadora para Avaliação de Cursos	1
Desenvolvimento de Desenho Industrial	-2
Conclusão de Especialização	-2
Mídia Social, website ou Blog	-3
Produção Técnica - Produto Tecnológico	-3
Participação em Banca Julgadora para Livre Docência	-5
Novos Estágios	-7
Novas Atividades de Pesquisa e Desenvolvimento	-8
Orientação em Andamento de Graduação	-10
Orientação em Andamento de Iniciação Científica	-11
Artigo Aceito para Publicação	-18
Novas Atividades de Ensino	-23
Participação em Direção ou Administração	-23
Orientação em Andamento de Mestrado	-27
Outras Orientações em Andamento	-28
Novas Atividades de Serviço Técnico-Especializado	-53
Conclusão de Pós-Doutorado	-54
Conclusão de Aperfeiçoamento	-67
Conclusão de Curso de Curta Duração	-82
Novos Projetos de Pesquisa	-155
Outras Orientações Concluídas	-331
Participação em Banca de Graduação	-716

Fonte: Produção própria.

A primeira coisa que se observa é que o impacto pode ser muito maior que o esperado.

Para se ter uma compreensão de sua dimensão, somente na área da educação ainda seria necessário examinar, por exemplo, os 1.236 novos vínculos, em sua maioria em atividades de professorado, 413 materiais didáticos produzidos, cada um com seu valor intrínseco que poderia ser calculado, além da enorme quantidade de materiais, inclusive editoriais, produzidos e sua relação com a educação. O impacto indireto na educação pode facilmente ultrapassar o próprio valor do financiamento e esse ainda se somaria ao impacto direto, na forma de produção de conhecimento e que pode ser verificado na forma de 6071 artigos publicados.

Uma análise rápida também pode indicar que a política pública de financiamento à pesquisa levou os pesquisadores a se concentrar em atividades mais especializadas e menos em atividades mais rápidas ou de menor “nível”. Um exemplo é o impacto negativo de 716 participações em banca de graduação, com o aumento de 1119 participações em banca de doutorado.

Um indicador que pode ser analisado com especial atenção é o impacto negativo em atividades de direção e administração, que representa um importante eixo da Tripla Hélice. Embora o impacto negativo tenha sido, na verdade, pequeno (23 pesquisadores, dos 13.104 financiados, teriam deixado alguma atividade de direção e administração), mostra-se que cabe estudo para verificar se ele é real e seu motivo.

Precisa-se lembrar que, como foram realizadas 33.600 regressões lineares, não é possível afirmar que o impacto representado na Tabela 2 está correto. Isso porque, assim como o impacto possui altas chances de estar seriamente subestimado, há uma alta chance estatística de ter sido listado como impacto algum erro estatístico, uma vez que o Grau de Confiança é de 99%. Assim, reforçamos a necessidade de um estudo detalhado para confirmar a veracidade do impacto listado antes de qualquer decisão a ser tomada com base nos achados.

Apesar disso, podemos inferir que o impacto do investimento realizado é alto, superando em muitas ordens o financiamento realizado, devendo, portanto, ser uma Política Pública não só a ser mantida, mas incentivada e expandida.