



**PROJETO: PESQUISA E ANÁLISE DE DADOS VINCULADOS AO
CAMPO DA SEGURANÇA PÚBLICA E SISTEMA PENITENCIÁRIO**

TERMO DE PARCERIA N° 817052/2015

Meta 05 – Etapa 01: Estudo técnico acerca das correlações entre os dados de ocorrências criminais e os do SIM/Datasus.

Versão atualizada em 15 de março de 2017.

(Primeira entrega realizada em 05 de julho de 2016)

FICHA INSTITUCIONAL

Conselho de Administração

Humberto de Azevedo Viana Filho
(Presidente)
Elizabeth Leeds – (Presidente de Honra)
Renato Sérgio de Lima (Vice Presidente)
Cássio Thyone Almeida de Rosa
Cristiane do Socorro Loureiro Lima
Haydée Glória Cruz Caruso
Jacqueline de Oliveira Muniz
Luciene Magalhães de Albuquerque
Marcos Aurelio Veloso e Silva
Marlene Ines Spaniol
Luis Flavio Saporì
Rodrigo Ghiringhelli de Azevedo
Silvia Ramos de Souza

Conselho Fiscal

Edson M. L. S. Ramos
Sérgio Roberto de Abreu

Equipe Executiva

Vice Presidente/Relações Institucionais
Renato Sérgio de Lima

Diretora Executiva
Samira Bueno

Coordenação Institucional

Patrícia Nogueira Pröglhöf

Coordenação de Projetos

Olaya Hanashiro

Equipe Técnica

Cauê Martins
David Marques
Marina Santos
Roberta Astolfi
Stefanie Prandi

Equipe Administrativa

Amanda Gouvêa
Débora Lopes
Karina Nascimento
Sueli Bueno

FICHA TÉCNICA DO TERMO DE PARCERIA N° 817052/2015

Supervisão Geral

Renato Sérgio de Lima

Coordenação do Projeto

Túlio Kahn

Coordenação Técnica

Olaya Hanashiro

Pesquisadoras

Samira Bueno

Patrícia Nogueira Pröglhöf

Roberta Astolfi

Marina dos Santos (Estagiária)

Stefanie Prandi (Estagiária)

Assistência Financeira e Administrativa

Débora Lopes

Auxílio Administrativo

Karina Nascimento

APRESENTAÇÃO

Este relatório se insere no projeto: Pesquisa e análise de dados vinculados ao campo da segurança pública e sistema penitenciário, referente ao termo de parceria N° 817052/2015 estabelecido entre o Ministério da Justiça e o Fórum Brasileiro de Segurança Pública. A meta 05 desse projeto compreende um estudo técnico acerca das correlações entre os dados de ocorrências criminais e os do Sistema de Informações de Mortalidade, SIM/Datasus, com o objetivo de verificar a validade dos dados de estatísticas criminais apresentados pelo Sistema Nacional de Informações de Segurança Pública - Sinesp. Além dessas correlações previstas, outras foram incluídas completando assim uma série de estudos a respeito de um mesmo tema dentro deste termo de parceria, que se distribui da seguinte forma:

- Meta 03, Etapa 01: Levantamento de iniciativas internacionais e no âmbito federal brasileiro para a padronização do registro de mortes violentas.
- Meta 03, Etapa 02: Estudo técnico mostrando as diferenças entre as categorias de dados de ocorrências criminais SIM/Datasus e a empregabilidade de cada uma de acordo com aspectos contextuais
- Meta 03, Etapa 03: Estudo técnico para atualização do quadro de equivalência (“de – para”) das mortes violentas dos sistemas estaduais de registro de ocorrências criminais.
- Meta 05: Estudo técnico das correlações entre os dados de ocorrências criminais e os do SIM/Datasus (o presente documento).

Este relatório está dividido em cinco seções principais, mais a lista de referências bibliográficas. Na introdução, discute-se a correlação ou ordem de grandeza entre os dados do Sinesp e do Datasus para Estados e municípios. Na segunda seção analisa-se a matriz de correlação de cada categoria de morte, nas duas bases, para o ano de 2014. A seguir, na terceira seção, analisa-se as categorias “mortes por causa mal definida” do Datasus e “mortes a esclarecer” do Sinesp. Na quarta seção, são apresentados três exercícios teóricos de validação para os dados do Sinesp. O primeiro exercício utiliza um grupo de variáveis externas à base de dados (variáveis socioeconômicas e demográficas); o segundo utiliza uma variável interna à base do Sinesp (ocorrências de porte ilegal de arma); e o terceiro exercício faz uma análise da distribuição dos casos de letalidade policial por Estado. Finalmente, na última seção são apresentadas algumas considerações sobre as diferentes metodologias que podem ser utilizadas para se verificar a qualidade de uma base de dados.

LISTA DE ABREVIATURAS E SIGLAS

Atlas PNUD – Atlas do Programa das Nações Unidas para o Desenvolvimento
Crisp - Centro de Estudos de Criminalidade e Segurança Pública
DEPEN – Departamento de Execução Penal
FNSP - Fundo Nacional de Segurança Pública
FBSP – Fórum Brasileiro de Segurança Pública
IBGE – Instituto Brasileiro de Geografia e Estatística
IDH – Índice de Desenvolvimento Humano
MDIP - Mortes em Decorrência de Intervenção Policial
MJ – Ministério da Justiça
MQO – Mínimos quadrados ordinários
OLS – *Ordinary least squares*
PAF – Projéteis de Armas de Fogo
PM – Polícia Militar
PNUD - Programa das Nações Unidas para o Desenvolvimento
Senasp – Secretaria Nacional de Segurança Pública
SIM/ Datasus – Sistema de Informação sobre Mortalidade
Sinesp - Sistema Nacional de Informações de Segurança Pública
SSP-SP – Secretaria de Estado da Segurança Pública
SUS – Sistema Único de Saúde

LISTA DE GRÁFICOS

Gráfico 1 - Correlação entre a taxa de agressões do Datasus e a taxa de homicídios Sinesp em 2014, por município.....	9
Gráfico 2 - Evolução temporal das séries de taxa de homicídios dolosos e taxa de porte ilegal de armas do Sinesp para os Estados de AL, BA, CE, RJ, RS e TO. Período de 3 meses.....	25
Gráfico 3 - correlação cruzada entre as séries de taxa de homicídios dolosos e taxa de porte ilegal de armas do Sinesp.....	26
Gráfico 4 - Gráfico de dispersão para a relação entre apreensão de armas e homicídios, por região.....	27
Gráfico 5 - Porcentagem de suicídio PAF (2013) e a taxa de apreensão de armas pela polícia por 100 mil habitantes (média de 2012 a 2014).	29
Gráfico 6 - Gráfico de dispersão: porcentagem de suicídio PAF (2013) e a taxa de apreensão de armas pela polícia por 100 mil habitantes (média de 2012 a 2014).	30
Gráfico 7 - Taxa de letalidade policial segundo o SUS e a SSP-SP no estado de São Paulo.	31
Gráfico 8 - Mapa da taxa média de MDIP em 2013-2014, por Estado.....	32
Gráfico 9 - Gráfico de dispersão: Letalidade policial observada X Letalidade predita pelo modelo	40

LISTA DE TABELAS

Tabela 1 - Dados de homicídios por agressões contabilizados pelo Datasus e a quantidade de mortes contabilizadas pelo Sinesp em 2014, por Estado e razão entre elas.....	8
Tabela 2 - Coeficientes das correlações dos números de homicídios Datasus e Sinesp, para os municípios em cada UF	10
Tabela 3 - Matriz de correlação de cada categoria de morte, nas bases do Datasus e do Sinesp para o ano de 2014	12
Tabela 4 - Porcentagem de óbitos por causas mal definidas – CID10,por ano segundo Unidade da Federação. Período: 1990-2011.....	13
Tabela 5- % de mortes a esclarecer no total de mortes no Sinesp, por UF. Período 2012 a 2014.	15
Tabela 6 - Modelo de regressão linear bayesiana para previsão da taxa de homicídios do SIM/Datasus em 2013, por mesorregião.	20
Tabela 7 - Variáveis utilizadas no modelo OLS (ou MQO).....	21
Tabela 8 Coeficientes R para as correlações entre apreensão de armas e homicídios, por região	27
Tabela 9 - Média de homicídios dolosos, a média de MDIP, porcentagem de MDIP no total de mortes e taxa de letalidade policial (todos dados do Sinesp). Na última coluna, taxa de letalidade policial segundo a categoria intervenção legal do Datasus. Período 2013-2014.	33
Tabela 10 - Variáveis preditoras utilizadas no modelo e suas fontes.....	34
Tabela 11 - Resumo do modelo.....	37
Tabela 12 - Teste ANOVA	37
Tabela 13 - Coeficientes	37
Tabela 14 - Taxa de letalidade policial	40

SUMÁRIO

1. Introdução	7
2. Matriz de Correlações SIM/Datasus - SINESP	11
3. Homicídios por causas mal definidas e mortes a esclarecer.....	13
4. Alternativas de validação para a base Sinesp	17
4.1. Validade externa: correlacionando os homicídios com variáveis socioeconômicas.....	18
4.2. Validade interna: correlacionando homicídios com armas de fogo	24
4.3. Letalidade Policial: avaliando políticas públicas.....	31
5. Considerações finais.....	42
6. Referência bibliográfica.....	43

1. INTRODUÇÃO

O objetivo deste produto é tanto verificar a consistência dos dados de homicídio coletados pelo Sinesp quanto ilustrar através de três ensaios originais as potencialidades de uso da base de dados para fins de conhecimento do fenômeno e formulação de políticas públicas.

No relatório da meta 3, etapa 2 já mostramos como há uma congruência entre os dados de mortes por agressões coletados pelo SIM/Datasus e os dados de homicídios e outras ocorrências com mortes coletados pelo Sinesp. Trata-se de um modo clássico de validação de dados através de fontes diferentes que medem fenômenos similares. O próprio SIM/Datasus lança mão deste procedimento de validação quando compara seus dados de morbidade com as projeções feitas pelo IBGE para a mortalidade nos municípios, levando em conta a idade e gênero da população.

Na tabela a seguir são apresentados os dados de homicídios por agressões contabilizados pelo Datasus e a quantidade de mortes contabilizadas pelo Sinesp (somando as categorias homicídio doloso, roubo seguido de morte, lesão corporal seguida de morte e outros crimes resultantes em morte) para o ano de 2014 e por Estado. Na última coluna, calculamos a razão entre as duas fontes, cuja média é de 1.14. Em outras palavras, a mortes contabilizadas pelo sistema de saúde são em média 14% superiores as mortes contabilizadas pelo sistema de segurança, o que é esperado.

Tabela 1 - Dados de homicídios por agressões contabilizados pelo Datasus e a quantidade de mortes contabilizadas pelo Sinesp em 2014, por Estado e razão entre elas.

Sigla UF	Agressões	PC-Qtde Ocorrências	Razão SUS / Sinesp
RR	154	67	2,30
PR	2829	1421	1,99
AP	245	148	1,66
RN	1500	940	1,60
AL	2093	1330	1,57
MS	688	491	1,40
DF	944	727	1,29
SP	5700	4496	1,27
AM	1226	978	1,25
SC	854	704	1,21
MG	4664	4066	1,15
MA	2351	2113	1,11
AC	232	211	1,10
RS	2660	2446	1,09
TO	362	336	1,08
GO	2782	2584	1,08
PB	1533	1434	1,07
RJ	5277	4949	1,07
SE	1096	1057	1,04
ES	1607	1559	1,03
CE	4408	4338	1,02
PI	714	710	1,01
PA	3322	3312	1,00
PE	3314	3337	0,99
MT	1330	1370	0,97
BA	5613	5966	0,94
RO	486	529	0,92
	57984	51585	1,12

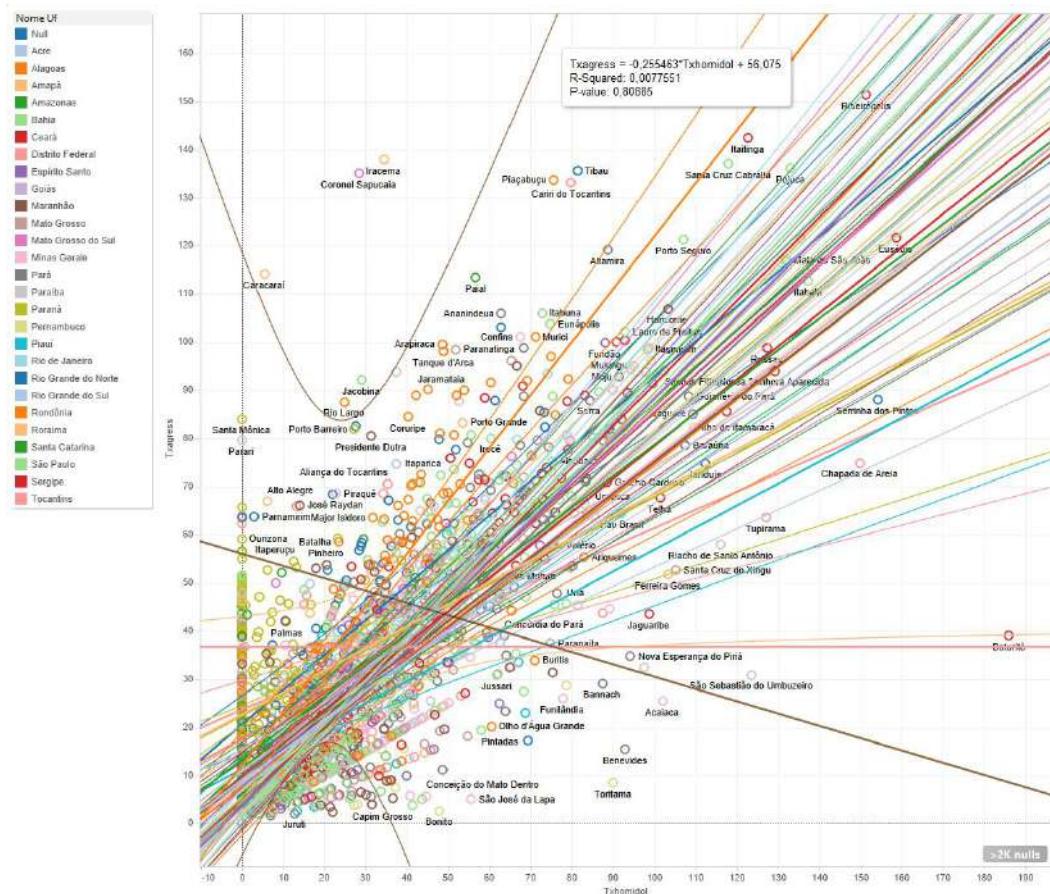
Fonte: Elaboração própria com dados do Datasus: Estatísticas vitais, mortalidade por causas externas, categorias CID-10 X85 - Y05 (disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937>) e Sinesp Sistema Nacional de Informações de Segurança Pública. Relatório Consolidado de Ocorrências de homicídios dolosos registrados pela Polícia Civil. Disponível em: <https://www.sinesp.gov.br/estatisticas-publicas>.

Vimos que nacionalmente a correlação ou ordem de grandeza entre as duas fontes é elevada, mas quando descemos ao nível de Estados ou cidades esta congruência diminui.

Nos demais Estados as diferenças são bem menores e em alguns deles, como PE, MT, BA e RO, os dados da segurança pública chegam a ser superiores, em números absolutos, aos dados da saúde. Se adotarmos como critério de qualidade uma diferença percentual de não mais de 10% entre as duas fontes, estariam em 2014 fora do padrão de qualidade 11 dos 27 Estados brasileiros: o RR, PR, AP, RN, AL, MS, SP, AM, SC, MG e MA.

O gráfico de dispersão abaixo mostra a correlação entre a taxa de agressões do Datasus e a taxa de homicídios Sinesp em 2014, por município. Os Estados são identificados pelas cores e as retas estão ajustadas por Estado, mostrando que em quase todos os Estados a correlação entre as duas medidas é forte e positiva.

Gráfico 1 - Correlação entre a taxa de agressões do Datasus e a taxa de homicídios Sinesp em 2014, por município.



Fonte: Elaboração própria com dados do Datasus: Estatísticas vitais, mortalidade por causas externas, categorias CID-10 X85 - Y05 (disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937>) e Sinesp planilha fornecida pela Coordenação Geral de Pesquisa e Analise da Informação em Segurança Publica - DEPAID/ SENASP/ MJ em 28 de maio de 2015).

A tabela abaixo descreve a equação da reta ajustada para cada Estado, com exceção do DF, para o qual só existe uma cidade. As equações sugerem que é possível prever adequadamente a taxa de homicídios do município tomando a taxa de agressões do Datasus. Apenas em Roraima o sinal foi negativo, provavelmente em função do tamanho pequeno da amostra de cidades (n=8). Também no Amapá (n=10) vemos uma correlação um pouco mais fraca entre as duas medidas, mas ainda assim, significativa a .004. Nos demais Estados, vemos que a correlação entre agressões Datasus e homicídios Sinesp é elevada, sugerindo que medem aproximadamente a mesma coisa.

Tabela 2 - Coeficientes das correlações dos números de homicídios Datasus e Sinesp, para os municípios em cada UF

Color	Line	Coefficients					
Nome_Uf	p-value	DF	Term	Value	StdErr	t-value	p-value
Tocantins	< 0,0001	75	Txhomidol	0,377265	0,08097	4,65931	< 0,0001
Sergipe	< 0,0001	64	Txhomidol	0,866378	0,0487464	17,7732	< 0,0001
São Paulo	< 0,0001	354	Txhomidol	0,751047	0,0346931	21,6483	< 0,0001
Santa Catarina	< 0,0001	139	Txhomidol	0,760652	0,0956929	7,94889	< 0,0001
Roraima	0,80885	8	Txhomidol	-0,255463	1,02164	-0,250051	0,80885
Rondônia	< 0,0001	25	Txhomidol	0,670452	0,110983	6,04105	< 0,0001
Rio Grande do Sul	< 0,0001	193	Txhomidol	0,871794	0,0406353	21,4541	< 0,0001
Rio Grande do Norte	< 0,0001	87	Txhomidol	0,723624	0,074295	9,73987	< 0,0001
Rio de Janeiro	< 0,0001	81	Txhomidol	0,978809	0,0442012	22,1444	< 0,0001
Piauí	< 0,0001	78	Txhomidol	0,456021	0,0707079	6,44937	< 0,0001
Pernambuco	< 0,0001	168	Txhomidol	0,764409	0,042148	18,1363	< 0,0001
Paraná	< 0,0001	255	Txhomidol	0,454842	0,0917915	4,95516	< 0,0001
Paraíba	< 0,0001	145	Txhomidol	0,621913	0,0476911	13,0404	< 0,0001
Pará	< 0,0001	122	Txhomidol	0,797611	0,0636122	12,5386	< 0,0001
Minas Gerais	< 0,0001	453	Txhomidol	0,674595	0,029617	22,7773	< 0,0001
Mato Grosso do Sul	< 0,0001	64	Txhomidol	0,75829	0,137343	5,52114	< 0,0001
Mato Grosso	< 0,0001	91	Txhomidol	0,688922	0,0728896	9,45158	< 0,0001
Maranhão	< 0,0001	127	Txhomidol	0,662933	0,069954	9,4767	< 0,0001
Espírito Santo	< 0,0001	66	Txhomidol	0,846182	0,0487502	17,3575	< 0,0001
Distrito Federal	N/A	0	Txhomidol	0	Because the trend line model response variable is constant, there is no information to estimate model statistics.		
Ceará	< 0,0001	99	Txhomidol	0,687636	0,0507329	13,5541	< 0,0001
Bahia	< 0,0001	308	Txhomidol	0,911089	0,0273943	33,2583	< 0,0001
Amazonas	< 0,0001	44	Txhomidol	0,654079	0,0821256	7,96437	< 0,0001
Amapá	0,0467433	10	Txhomidol	0,458427	0,202144	2,26782	0,0467433
Alagoas	< 0,0001	99	Txhomidol	1,09886	0,0900166	12,2073	< 0,0001
Acre	< 0,0001	15	Txhomidol	0,620768	0,106745	5,81541	< 0,0001

Fonte: Elaboração própria com dados do Datasus: Estatísticas vitais, mortalidade por causas externas, categorias CID-10 X85 - Y05 (disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937>) e Sinesp planilha fornecida pela Coordenação Geral de Pesquisa e Analise da Informação em Segurança Publica - DEPAID/ SENASP/ MJ em 28 de maio de 2015).

2. MATRIZ DE CORRELAÇÕES SIM/DATASUS - SINESP

Até agora falamos de contagem de mortes, indistintamente, como se tivéssemos lidando com fenômenos unidimensionais. Quando calculamos a taxa de homicídios do Sinesp, somamos os casos de homicídios aos de lesões corporais seguidas de morte, roubos seguidos de morte e outros crimes seguidos de morte, de modo a obter uma quantidade de casos mais próxima a do Datasus, que não faz estas distinções. Mas na verdade sabemos que os fenômenos mensurados em cada categoria tratam de coisas bastante diferentes. Um homicídio passional tem pouca ligação com um latrocínio, cuja motivação é patrimonial e estes guardam tênue relação com as mortes em decorrência de intervenção policial. Um modo de investigar se estamos falando de fenômenos uni ou multidimensionais é através da matriz de correlação entre as variáveis.

Por definição, sabemos que um homicídio passional é diferente de um latrocínio e estes das intervenções legais. Mas como se comportam categorias mais ambíguas, como “eventos cuja intenção é indeterminada” ou “mortes a esclarecer”? O fato de eles acontecerem nos mesmos locais que outros tipos específicos de mortes nos diz algo sobre sua real natureza? Tratar-se-iam, por exemplo, de homicídios camuflados nas estatísticas? A Matriz de correlações usa um indicador de semelhança – o coeficiente de Pearson – para dar pistas sobre a natureza destas categorias.

Neste tópico apresentamos a matriz de correlação de cada categoria de morte, nas duas bases, para o ano de 2014. As bases foram mescladas usando o nome da cidade, com os problemas já apontados no relatório anterior (meta 03, etapa 03). Quando tomamos os números absolutos, vemos que basicamente “tudo de correlaciona com tudo” e praticamente todos os coeficientes são significativos: um município que tem mais mortes de um tipo tende também a ter mais mortes dos outros tipos. Na maioria dos casos estamos apenas medindo o tamanho das cidades, pois o risco de morte é uma função do tamanho da população.

A análise fica mais interessante quando observamos as taxas por 100 mil habitantes, controlando, portanto, o fator tamanho de população.

Como esperado, a taxa de mortes por agressões do Datasus no município se correlaciona fortemente com a taxa de homicídios do Sinesp ($r= .75$) e em medida bem menor com as lesões corporais seguidas de morte e roubos seguidos de morte.

Os eventos cuja intenção é indeterminada, do Datasus, por sua vez, não apresentam correlações dignas de menção com nenhum outro indicador. Esta pode ser uma evidência de que ela não oculta muitos homicídios, embora a literatura já tenha identificado que muitos deles são de fato homicídios, quando os casos são reanalisados por outros meios. (CERQUEIRA, Daniel. Mortes violentas não esclarecidas e impunidade no Rio de Janeiro. Econ. Apl. [online]. 2012, vol.16, n.2 [cited 2016-07-05], pp.201-235). O mesmo pode ser dito com relação à taxa de mortes a esclarecer do Sinesp: a correlação com homicídios dolosos no município é significativa estatisticamente, porém pequena. Se ela “esconde” homicídios dolosos, são provavelmente poucos.

Tabela 3 - Matriz de correlação de cada categoria de morte, nas bases do Datasus e do Sinesp para o ano de 2014 .

		Correlations							
		txagress	txeventos	txintervencoes	txhomidol	txescorp	txesclarecer	txoutros	txlatro
txagress	Pearson Correlation	1	-.022	-.060	.750**	.101**	.013	.038	.094**
	Sig. (2-tailed)		.453	.562	.000	.000	.530	.068	.000
	N	3221	1186	96	3221	2865	2410	2286	3221
txeventos	Pearson Correlation	-.022	1	.177	-.009	-.023	-.073*	-.063	.001
	Sig. (2-tailed)	.453		.138	.736	.439	.029	.070	.965
	N	1186	1408	72	1408	1180	900	837	1408
txintervencoes	Pearson Correlation	-.060	.177	1	-.101	.083	-.085	-.183	-.100
	Sig. (2-tailed)	.562	.138		.316	.458	.476	.133	.320
	N	96	72	101	101	82	73	69	101
txhomidol	Pearson Correlation	.750**	-.009	-.101	1	.094**	.107**	.046*	.087**
	Sig. (2-tailed)	.000	.736	.316		.000	.000	.010	.000
	N	3221	1408	101	4785	4158	3311	3193	4785
txescorp	Pearson Correlation	.101**	-.023	.083	.094**	1	-.009	.005	.047**
	Sig. (2-tailed)	.000	.439	.458	.000		.584	.768	.002
	N	2865	1180	82	4158	4240	3348	3241	4222
txesclarecer	Pearson Correlation	.013	-.073*	-.085	.107**	-.009	1	.047**	.014
	Sig. (2-tailed)	.530	.029	.476	.000	.584		.008	.412
	N	2410	900	73	3311	3348	3358	3219	3347
txoutros	Pearson Correlation	.038	-.063	-.183	.046*	.005	.047**	1	.108**
	Sig. (2-tailed)	.068	.070	.133	.010	.788	.008		.000
	N	2286	837	69	3193	3241	3219	3254	3237
txlatro	Pearson Correlation	.094**	.001	-.100	.087**	.047**	.014	.108**	1
	Sig. (2-tailed)	.000	.965	.320	.000	.002	.412	.000	
	N	3221	1408	101	4785	4222	3347	3237	4862

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Fonte: Elaboração própria com dados do Datasus: Estatísticas vitais, mortalidade por causas externas, categorias CID-10 X85 - Y05 (disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937>) e Sinesp planilha fornecida pela Coordenação Geral de Pesquisa e Analise da Informação em Segurança Publica - DEPAID/ SENASP/ MJ em 28 de maio de 2015).

A matriz confirma, portanto, a congruência entre as bases do Datasus e do Sinesp no que tange à mensuração dos homicídios e ao mesmo tempo revela que estamos diante de um fenômeno multidimensional. É preciso ter esta multidimensionalidade em mente quando formos analisar os experimentos preditivos mais a frente, pois o conjunto de preditores dos homicídios dolosos derivados de conflitos interpessoais pode ser completamente diferente dos preditores de mortes em intervenções legais ou latrocínios, embora sejam todas partes do conjunto de mortes violentas intencionais.

3. HOMICÍDIOS POR CAUSAS MAL DEFINIDAS E MORTES A ESCLARECER

Os profissionais que analisam os dados de mortalidade da saúde sugerem que a qualidade das estatísticas de mortes é razoavelmente adequada quando as mortes por causa mal definida ficam abaixo de 10% do total de mortes (MINISTÉRIO DA SAÚDE, 2008).

O sistema de coleta na área de saúde é mais antigo (existe desde 1979) e consolidado e permite avaliar a melhora da qualidade da informação nas últimas décadas. Os óbitos por causas mal definidas passaram de 18,2% para 6,7% do total de óbitos, uma melhora de 63% entre 1990 e 2011. Observe-se que em alguns Estados a situação se deteriorou, como no Amapá e no Rio de Janeiro, mas na maioria das unidades houve uma sensível melhora na coleta. Se adotarmos o critério de 10%, apenas AM, BA e PA seriam classificados como dados de “baixa qualidade” em 2011.

Tabela 4 - Porcentagem de óbitos por causas mal definidas – CID10, por ano segundo Unidade da Federação. Período: 1990-2011.

Unidade da Federação	1990	1995	2000	2005	2010	2011	Variação
Amazonas	35,40	25,60	25,20	21,30	15,30	13,80	-61,02
Bahia	32,80	30,30	29,70	25,30	13,30	13,50	-58,84
Pará	30,80	29,30	27,60	22,40	12,90	12,70	-58,77
Amapá	7,00	6,30	14,90	14,30	13,60	9,60	37,14
Acre	32,00	25,50	31,20	7,70	13,60	9,10	-71,56
Minas Gerais	16,40	15,00	14,20	11,50	9,90	9,00	-45,12
Rondônia	14,00	18,40	12,90	9,40	9,20	8,20	-41,43
Paraíba	55,70	50,20	46,50	15,60	7,90	7,70	-86,18
Alagoas	47,10	40,00	29,10	13,50	8,60	7,60	-83,86
Rio de Janeiro	5,60	9,80	11,50	9,30	7,40	6,80	21,43
Mato Grosso	16,30	15,20	7,90	6,40	4,10	6,60	-59,51
Sergipe	44,40	33,00	29,70	9,50	5,40	6,50	-85,36
Maranhão	49,50	42,40	36,10	16,60	6,70	6,40	-87,07
São Paulo	6,20	6,60	6,60	6,30	6,20	6,00	-3,23
Goiás	15,90	18,00	10,20	7,20	6,10	5,60	-64,78
Piauí	39,50	36,00	30,70	22,10	4,30	5,20	-86,84
Pernambuco	42,20	27,90	22,80	10,00	5,40	5,20	-87,68

Ceará	44,40	38,10	20,20	19,10	5,10	4,80	-89,19
Paraná	14,30	10,80	5,40	4,80	5,10	4,80	-66,43
Rio Grande do Sul	8,00	6,10	4,40	5,40	4,50	4,60	-42,50
Santa Catarina	17,30	15,10	12,60	8,70	5,30	4,00	-76,88
Rio Grande do Norte	37,30	30,30	27,60	8,50	3,00	3,60	-90,35
Roraima	8,90	8,30	9,10	6,20	4,60	3,40	-61,80
Tocantins	41,90	32,40	21,20	5,50	2,70	3,10	-92,60
Mato Grosso do Sul	11,40	10,40	8,50	1,80	1,40	1,70	-85,09
Espírito Santo	20,80	20,20	17,60	4,40	1,40	1,30	-93,75
Distrito Federal	2,10	1,50	4,70	2,30	1,00	1,10	-47,62
Total	18,20	16,20	14,30	10,40	7,00	6,70	-63,19

Fonte: Elaboração própria com dados do Datasus: Estatísticas vitais, mortalidade por causas externas, categorias CID-10 R95 – R99 (disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937>) e

Seria possível fazer algo semelhante com os dados de mortes da segurança? Uma alternativa para averiguar a qualidade das estatísticas de mortes na esfera da segurança, fazendo um paralelo com a saúde, seria computar a porcentagem de mortes a esclarecer em relação ao total de mortes. Conceitualmente estamos falando de coisas diferentes, mas em última instância ambos dependem de mais informações sobre a morte para aprimorar a classificação: na saúde, talvez a lacuna de informações exija mais exames ou averiguação dos relatos policiais e, na segurança, de mais investigação, evidências testemunhais ou periciais.

Nos dois casos sempre haverá um resíduo de situações que não permitem a classificação segura, mas a experiência da Saúde mostra que é possível aprimorar consideravelmente a qualidade das informações. Embora conceitualmente diferentes, é digno de nota observar que quando analisamos os municípios, encontramos um R2 de 0.46 entre os “eventos cuja intenção é indeterminada” do Datasus e as “mortes a esclarecer” do Sinesp. Os médicos legistas se valem bastante dos dados da ocorrência para a classificação das mortes de modo que faz sentido pensar que onde existe uma grande porcentagem de mortes a esclarecer exista igualmente elevada porcentagem de eventos cuja intenção é indeterminada.

Na tabela a seguir estão somados os dados do período 2012 a 2014, para dar mais estabilidade aos números. As categorias foram somadas e, na última coluna, foi computada a porcentagem de mortes a esclarecer, pelo qual os Estados estão listados em ordem decrescente.

Tabela 5- % de mortes a esclarecer no total de mortes no Sinesp, por UF. Período 2012 a 2014.

UF	Homicídio doloso	Lesão corporal seguida de morte	Outros crimes resultantes em morte	Roubo seguido de morte (latrocínio)	Mortes a esclarecer	Total	% a esclarecer
SC	1.852	173	127	120	2.967	5.239	56,6%
MS	1.279	32	6	46	1435	2.798	51,3%
RJ	12.893	101	419	427	12.776	26.616	48,0%
RS	6.056	148	204	335	5986	12.729	47,0%
RR	226	10	0	9	195	440	44,3%
RO	1.436	18	49	46	923	2.472	37,3%
SE	2.682	24	11	122	1513	4.352	34,8%
MG	1.1977	198	46	279	5.666	18.166	31,2%
TO	896	21	8	42	392	1.359	28,8%
GO	2.804	8	34	107	901	3.854	23,4%
PE	9.135	228	14	235	2.812	12.424	22,6%
PR	4.038	262	39	131	1.258	5.728	22,0%
RN	2.084	193	27	40	604	2.948	20,5%
CE	11.776	205	346	275	3.221	15.823	20,4%
MT	3.210	156	66	133	782	4347	18,0%
PI	1.639	48	5	82	266	2.040	13,0%
AC	562	7	6	42	87	704	12,4%
ES	4.753	33	25	84	504	5.399	9,3%
BA	15.241	350	66	434	1.232	17.323	7,1%
MA	4.937	316	30	186	322	5.791	5,6%
AM	2.811	77	36	121	137	3.182	4,3%
PA	9.437	108	-	453	214	10.212	2,1%
AP	504	9	1	25	11	550	2,0%
PB	4.238	13	3	57	87	4.398	2,0%
DF	2.161	13	0	119	37	2.330	1,6%
AL	5.252	52	15	178	19	5.516	0,3%
SP	13.574	-	-	1.096		14.670	0,0%
	123.879	2.803	1.583	4.128	44.347	176.740	25,1%

Fonte: Elaboração própria com dados do Sistema Nacional de Informações de Segurança Pública. Sinesp planilha fornecida pela Coordenação Geral de Pesquisa e Analise da Informação em Segurança Publica - DEPAID/ SENASP/MJ em 28 de maio de 2015).

Vários dados estão ausentes para São Paulo, que foi excluído do total. Vemos que em média, 25% das mortes coletadas pelo sistema de segurança estão ainda categorizadas como “mortes a esclarecer”, o que equivaleria para efeitos de comparação, às mortes por causas mal definidas na saúde. Em muitos Estados o percentual de mortes a esclarecer é bastante elevado e pode afetar consideravelmente as estatísticas de homicídios dolosos caso muitos deles seja de fato homicídios. Note-se que entre os Estados mais problemáticos estão SC, RJ e RS e no extremo oposto AL, DF, PB, AP, PA, AM, MA, e BA , UFs do Norte e Nordeste em sua maioria, que seriam classificadas como de “alta qualidade” caso o critério de 10% de mortes a esclarecer fosse adotado.

Vemos, portanto que a congruência entre as bases de dados do Sinesp e do Datasus é elevada, mas que existem também diferenças significativas quando descemos aos detalhes. Os critérios são em parte arbitrários, mas mostramos que em muitos Estados as diferenças entre as duas bases é superior a 10%. E o problema decerto é maior quando olhamos para cidades. Do mesmo modo, vimos que o percentual de “mortes a esclarecer” no total de mortes computadas pelo sistema de segurança é bastante elevado e que se adotássemos o critério dos 10%, a maioria dos Estados teria suas estatísticas qualificadas como de “baixa qualidade”.

4. ALTERNATIVAS DE VALIDAÇÃO PARA A BASE SINESP

No que diz respeito aos procedimentos de verificação de consistência dos dados, nos propomos a fazer nos tópicos seguintes alguns exercícios teóricos de validação, usando para isso variáveis que a literatura criminológica sugere estarem correlacionadas aos homicídios (BARATA, 1998; NERY, 2005; DOS SANTOS E KASSOUF, 2008; SOARES, BATITUCCI E RIBEIRO, 2007; ANDRADE, 2011; PERES et al., 2011; DUARTE, 2012; MANSANO, 2013; CORREIA, 2014; SOUSA, 2014; SOUZA, 2014). Primeiro lançando mão de um grupo de variáveis externas à base de dados, como um grupo de variáveis socioeconômicas e demográficas compiladas pelo Altas do PNUD e cuja fonte original é o IBGE.

Em seguida, usando uma variável que pertence à base de dados Sinesp, especificamente as ocorrências de porte ilegal de arma. A lógica subjacente é semelhante: comparar os dados observados com os dados previstos, como base na relação linear entre a distribuição dos homicídios e um conjunto de variáveis preditoras. Os exercícios devem, portanto, sendo consistentes os dados, corroborar estas correlações já identificadas na literatura. Além disso, novamente, eles são um procedimento para identificar casos desviantes e anômalos, que merecem ser investigados mais a fundo ou por se tratarem de erros e fraudes ou porque indicam eventualmente uma política bem-sucedida ou alguma característica típica relevante para explicar o desvio.

No terceiro exercício faremos uma análise inédita da distribuição dos casos de letalidade policial por Estado, procurando testar algumas variáveis que expliquem por que a letalidade está mais concentrada em alguns locais do que em outros. Os dados do sistema de justiça criminal sobre mortes em confronto são melhores que os compilados pelo Datasus (Y35-Y36: Intervenções legais e operações de guerra) e utilizaremos aqui variáveis provenientes majoritariamente da pesquisa IBGE Estadic, que poucas vezes foram tratadas na literatura, uma vez que a Estadic é relativamente recente.

Todos os exercícios aqui alinhavados são de natureza exploratória e cada um deles merece mais aprofundamentos. As conclusões são provisórias e valem como exemplos de como os dados podem ser trabalhados para responder a questões de interesse para os acadêmicos e gestores públicos.

Estes exercícios são também exemplos do que a comunidade acadêmica pode fazer e contribuir para o entendimento do fenômeno ou para a avaliação de políticas públicas, uma vez que os dados estejam abertos de forma mais ampla e disagregada possível. Os usuários dos dados abertos governamentais agregam valor aos mesmos. O MJ/Sinesp presta um enorme serviço à sociedade simplesmente coletando, padronizando e divulgando informações de interesse público.

4.1. Validade externa: correlacionando os homicídios com variáveis socioeconômicas

No relatório anterior, usamos duas fontes para homicídios – Datasus e Sinesp - que medem fenômenos similares para fazer uma checagem externa da consistência das informações. Mas nada impede que, por algum motivo, ambas estejam equivocadas, algumas vezes, quando consideramos localidades específicas. Frequentemente, os certificados de óbito usados como fonte pelo SUS são elaborados pelos médicos legistas que fazem parte da polícia científica, de modo que as fontes originárias são muitas vezes as mesmas, tanto para a saúde quanto para as SSPs. Como mencionado anteriormente, um procedimento de checagem alternativo é comparar os números observados de mortes (tanto pelo SUS quanto pelo Sinesp) com os números “esperados”, ou seja, aqueles previstos em função da elevada correlação dos homicídios com uma combinação linear de outras variáveis socioeconômicas ou demográficas.

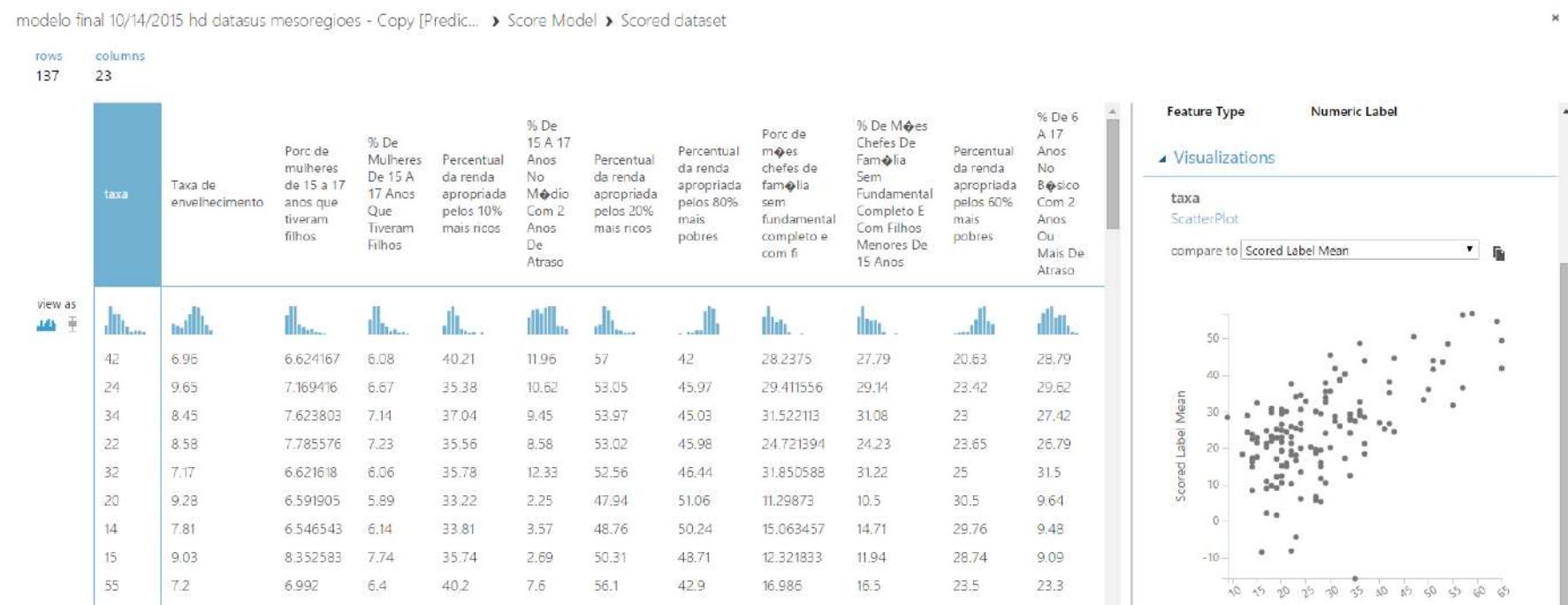
A literatura criminológica já apontou que os homicídios estão correlacionados com a disponibilidade de armas, grau de urbanização, proporção de jovens de sexo masculino na população, índice de desigualdade, taxa de desemprego, uso de álcool e drogas, porcentagem de residências uniparentais, etc. (ver relatório da meta 01, etapas 01 e 02).

É possível então lançar mão de procedimentos estatísticos, como análise de regressão, para produzir estimativas da quantidade de homicídios em cada cidade, com base no valor destas outras variáveis. Trata-se sempre de uma estimativa probabilística e é preciso ter em mente que estas estimativas são imperfeitas, pois as variáveis preditoras também estão sujeitas a erros de mensuração, coleta, digitação, etc. Ainda assim pode ser uma ferramenta útil para validação da qualidade da base ou identificação de casos excepcionais para estudo.

O modelo abaixo foi gerado na *machine learning* do Microsoft Azure e faz uma previsão da taxa de homicídios do SIM/Datasus em 2013, por mesorregião do país, em função de 22 variáveis socioeconômicas e demográficas. Utilizamos as 137 mesorregiões geográficas como unidade de análise, pois o modelo fica mais robusto do que quando utilizamos os mais de 5 mil municípios, pois a maioria deles tem poucos ou nenhum homicídio. O Azure permite testar diversos modelos de regressão, *on-line*, treinando a base conforme novas informações são agregadas. Obviamente que o sistema não resolve os problemas típicos encontrados nas regressões: multicolinearidade entre as variáveis, endogeneidade, heterocedasticidade, etc., encontrados em qualquer manual de regressão. *Data Science*, adotado como paradigma no mundo empresarial, é mais *Data* do que *Science* e não deve necessariamente ser replicado acriticamente pelas instituições públicas. A vantagem, por outro lado, é poder trabalhar com uma grande massa de dados e “falhar rápido”. O sistema permite testar simultaneamente diversos modelos e comparar seus resultados de modo a escolher o mais adequado. Uma *machine learning* deste tipo é bastante interessante para as polícias estaduais, onde novos dados de crimes são enviados a base operacional a cada segundo. Estes novos dados retroalimentam o sistema, checando se os parâmetros são adequados ou se devem ser modificados. É nesta característica que reside o “aprendizado de máquina”, em sua capacidade de melhorar a previsão continuamente, enquanto milhares de novas informações são atualizadas.

O modelo de regressão linear bayesiana prevê razoavelmente bem a taxa de homicídios na mesorregião ($r=.61$). Na tabela a seguir, o Azure mostra uma parte da lista de 23 variáveis que mais contribuem para explicar os homicídios: taxa de envelhecimento, % de mulheres de 15 a 17 anos que tiveram filhos, porcentagem da renda apropriada pelos 10% mais ricos, e assim por diante, por ordem de relevância.

Tabela 6 - Modelo de regressão linear bayesiana para previsão da taxa de homicídios do SIM/Datasus em 2013, por mesorregião.



Fonte: Elaboração própria com dados do Datasus: Estatísticas vitais, mortalidade por causas externas, categorias CID-10 X85-Y09. Atlas do Desenvolvimento Humano no Brasil. Programa das Nações Unidas para o Desenvolvimento. Brasil, 2013.

O diagrama de dispersão do lado direito mostra a relação entre a taxa de homicídio observada e a taxa prevista pelo modelo. O modelo pode ser aperfeiçoado, mas não é este o ponto. Novamente aqui, evidencia-se que a distribuição dos homicídios pelo território não é aleatória e que os homicídios se relacionam, como esperado, com outras variáveis já conhecidas na literatura criminológica. É um atestado indireto de que a base mede com razoável fidedignidade o fenômeno que está sendo classificado como homicídio doloso. Ela permite fazer previsões, inferências, avaliações, estudar casos específicos de sucesso, identificar erros e tentativas de manipulação.

O uso de uma ferramenta *on-line* deste tipo pelo Ministério da Justiça só se justificaria para uma massa muito grande de dados, atualizada a pequenos intervalos. O Sinesp recebe dados mensalmente, por Distrito Policial, o que implica numa base pequena para os padrões atuais e, portanto, perfeitamente manipulável em um servidor munido de um *software* estatístico, mesmo gratuito, como no exemplo a seguir.

Neste segundo exemplo, utilizamos o *software* livre Geoda para rodar uma regressão pelo método dos Quadrados Mínimos Ordinários (MQO ou OLS do inglês Ordinary Least Squares) usando a taxa de homicídios de 2012 como variável dependente e uma série de variáveis preditoras de ordem social, econômica e institucionais. Trabalhamos agora apenas com 27 Unidades Federativas, ao invés das 137 mesorregiões, uma vez que algumas variáveis provenientes da pesquisa Estadic/IBGE só fazem sentido a este nível.

O ponto é que existem definições em nossos modelos que nenhum programa estatístico é capaz de decidir, mas apenas o pesquisador. No primeiro exemplo, optamos por não trabalhar em nível de municípios, pois como homicídio é um fenômeno raro, restariam muitas células com 0 casos, o que é ruim para a modelagem. Neste caso, optamos por agregar os dados por UF, simplesmente por que alguns dados estão disponíveis apenas a este nível de desagregação. O problema está em que a escolha da unidade de análise adequada tem suas consequências, pois o impacto de alguns fatores só se revela em certo nível de agregação do dado.

A abordagem para a escolha das variáveis também é diferente: ao invés de simplesmente adicionar um grande número de variáveis e deixar que o programa escolha os melhores preditores, numa abordagem tipo *data-driven*, comum na era do *BigData*, no modelo abaixo a escolha das variáveis foi ditada antes por razões teóricas. Queríamos explicitamente aqui testar diversas hipóteses, tais como quanto maior a despesa percentual do Estado com segurança pública, menor a taxa de homicídios ou a taxa de homicídios é menor nos Estados com Planos de Prevenção de homicídios ou com Ouvidorias, e assim por diante.

Tabela 7 - Variáveis utilizadas no modelo OLS (ou MQO).

Variável	Observação	Fonte
Part_Desp_SP_%	Despesas com Segurança Pública em %	FBSP
PRPH	Plano de redução e/ou prevenção de homicídios – Dummy (se possui, sim ou não)	Estadic/IBGE

		(2014)
Letalidade_Policial_TXP	(Letalidade policial 2013+2014)/2)/n. de policiais * 100000 - Variável Y - Dependente	FBSP
TX_PRHD	Taxa de prisão de homicidas	DEPEN
Média de txdes_ano	Média de taxa de desemprego	Microdados IBGE
Suici_Arma_%	% de suicídios por arma de fogo. Proxy para armas	Datasus
Avg. % de crianças extremamente pobres	% média de crianças extremamente pobres 2010	Atlas PNUD
Ouvidoria	Possui ouvidoria - Dummy (sim ou não)	Estadic/IBGE (2014)

O modelo tem um ajuste geral bastante bom ($R^2 = .82$) e, com exceção de existência de Ouvidoria, todas as demais variáveis foram significativas, como pode ser checado mais adiante.

A regressão sugere que quanto maior o percentual de despesa do Estado em segurança, maior a taxa de homicídios dolosos. Há aqui provavelmente um problema de simultaneidade: provavelmente, os Estados que viram os homicídios crescer nos últimos anos terminaram por investir fatias crescentes do orçamento em segurança, para tentar contornar o problema. Isto explicaria a associação positiva encontrada entre gastos e homicídios, que do contrário faz pouco sentido.

Como esperado, os Estados que afirmaram ter plano de redução ou prevenção de homicídios têm realmente menores taxas de homicídio que os demais, o que sugere que dar prioridade política ao tema pode ter impactos significativos sobre a criminalidade.

Conforme veremos em detalhe mais a frente, existe uma relação positiva entre a taxa de homicídios e a taxa de letalidade policial: é possível que o medo dos policiais que atuam em áreas violentas estimule os confrontos ou que haja uma permissividade maior para os confrontos em locais assolados pelos homicídios. Qualquer que seja o motivo, a relação entre homicídios em geral e confrontos letais é positiva e significativa.

A priori, a taxa de prisão de homicidas é uma medida de impunidade e esperaríamos encontrar uma relação inversa entre a prisão de homicidas e a taxa de homicídios, na medida em que a prisão funcione como fator dissuasor. O sinal encontrado, todavia, foi positivo. Novamente aqui, é possível que estejamos diante de um caso de causalidade reversa: estados com muitos homicídios estão prendendo mais homicidas, como parte da estratégia de combate ao crime, o que explicaria a relação positiva no curto prazo entre as duas variáveis. Esta é uma das limitações do design *cross-section*, que analisa apenas as diferenças entre as unidades territoriais num momento do tempo. Um *design* de dados em painéis temporais e o uso de variáveis instrumentais para lidar com a questão da simultaneidade são necessários para estabelecer a existência e sinal de uma relação.

A literatura que analisou o tema já identificou, na maioria das vezes, a associação positiva entre homicídios e as condições socioeconômicas locais. Este experimento corrobora esta

associação, pois as taxas de homicídio revelaram-se maiores nos estados que tiveram maiores taxas de desemprego, bem como maior percentual médio de crianças extremamente pobres.

Novamente corroborando o identificado na literatura recente, encontramos uma correlação positiva e significativa entre taxa de homicídios e a disponibilidade de armas de fogo, medida aqui, como já usual, por sua substituta: a porcentagem de suicídios cometidos com arma de fogo no total de suicídios.

Da pesquisa Estadic/IBGE usamos a informação da existência ou não de Ouvidoria de polícia no Estado. Achamos interessante reportar o sinal negativo entre homicídios e existência de Ouvidoria, que é o que se espera. Todavia, como notado anteriormente, a correlação não é estatisticamente significativa.

SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION

R-squared	:	0.821933	F-statistic	:	10.3857
Adjusted R-squared	:	0.742792	Prob(F-statistic)	:	2.3474e-005
<hr/>					
Variable	Coefficient	Std.Error	t-Statistic	Probability	
CONSTANT	-30.06139	8.693958	-3.457734	0.00281	
Part_Desp_SP_%	2.265319	0.4224205	5.362711	0.00004	
PRPH	-10.47928	3.003312	-3.489242	0.00262	
Letalidade_Policial_TXP	0.02202537	0.005719491	3.850933	0.00117	
TX_PRHD	0.220067	0.0502249	4.381632	0.00036	
Média de txdes_ano	1.079238	0.3896699	2.769621	0.01263	
Suici_Arma_%	0.8315747	0.3623147	2.295172	0.03396	
% de crianças extr. pobres	0.406872	0.09942309	4.092329	0.00068	
Ouvidoria	-4.054615	3.019242	-1.342925	0.19599	

REGRESSION DIAGNOSTICS

MULTICOLLINEARITY CONDITION NUMBER 23.069719

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	1.1371	0.56633

DIAGNOSTICS FOR HETEROSKEDASTICITY

RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	8	7.6226	0.47118
Koenker-Bassett test	8	8.8941	0.35130

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	44	27.0000	0.97955

OBS	Avg. taxa hd 2012 anual	PREDICTED	RESIDUAL
1	25.50000	32.35917	-6.85917
2	58.86000	56.22423	2.63577
3	23.40000	18.74021	4.65979
4	26.42000	20.88270	5.53730
5	34.45000	35.41661	-0.96661
6	47.16000	39.86592	7.29408
7	24.80000	27.20744	-2.40744
8	40.75000	30.70057	10.04943
9	29.05000	30.02791	-0.97791
10	23.21000	24.59110	-1.38110
11	19.89000	20.76075	-0.87075
12	17.35000	19.10724	-1.75724
13	31.46000	29.22028	2.23972
14	39.69000	41.43261	-1.74261
15	37.46000	34.10951	3.35049
16	30.94000	31.69779	-0.75779

17	15.73000	17.55595	-1.82595
18	11.18000	17.53244	-6.35244
19	27.01000	24.32331	2.68669
20	22.58000	30.24042	-7.66042
21	26.91000	34.55727	-7.64727
22	20.28000	23.88104	-3.60104
23	16.51000	5.79781	10.71219
24	10.02000	13.70648	-3.68648
25	40.08000	36.62259	3.45741
26	10.18000	11.67857	-1.49857
27	19.42000	22.05007	-2.63007

Fonte: Elaboração própria a partir dos dados e fontes referidos na tabela sete.

Vimos aqui que o pesquisador precisa tomar uma série de decisões importantes antes de simplesmente rodar seu modelo. Embora melhor do que uma análise bivariada, uma análise de regressão não resolve todos os problemas, tais como endogeneidade e simultaneidade entre as variáveis. Às vezes é preciso que a coleta de dados se estenda por vários anos em todos os Estados, de modo a permitir uma análise em desenho de painel, mais adequada para resolver questões como a do “sinal” de uma relação. Correlação, como é comum alertar, não é “causação”.

4.2. Validade interna: correlacionando homicídios com armas de fogo

Vimos acima dois procedimentos para checar a qualidade dos dados de homicídios na esfera da segurança pública, usando dados por mesorregiões e Estados: correlacionar com a base com outras variáveis externas tradicionalmente ligadas aos homicídios. Os procedimentos nos ajudam a identificar eventuais erros, casos de interesse e, eventualmente, fatores de risco e proteção por traz do fenômeno. Um *outlier* pode ser um erro, uma fraude ou um grande *case* de sucesso para ser estudado e disseminado! Estes casos excepcionais só podem ser identificados quando comparados a algum padrão.

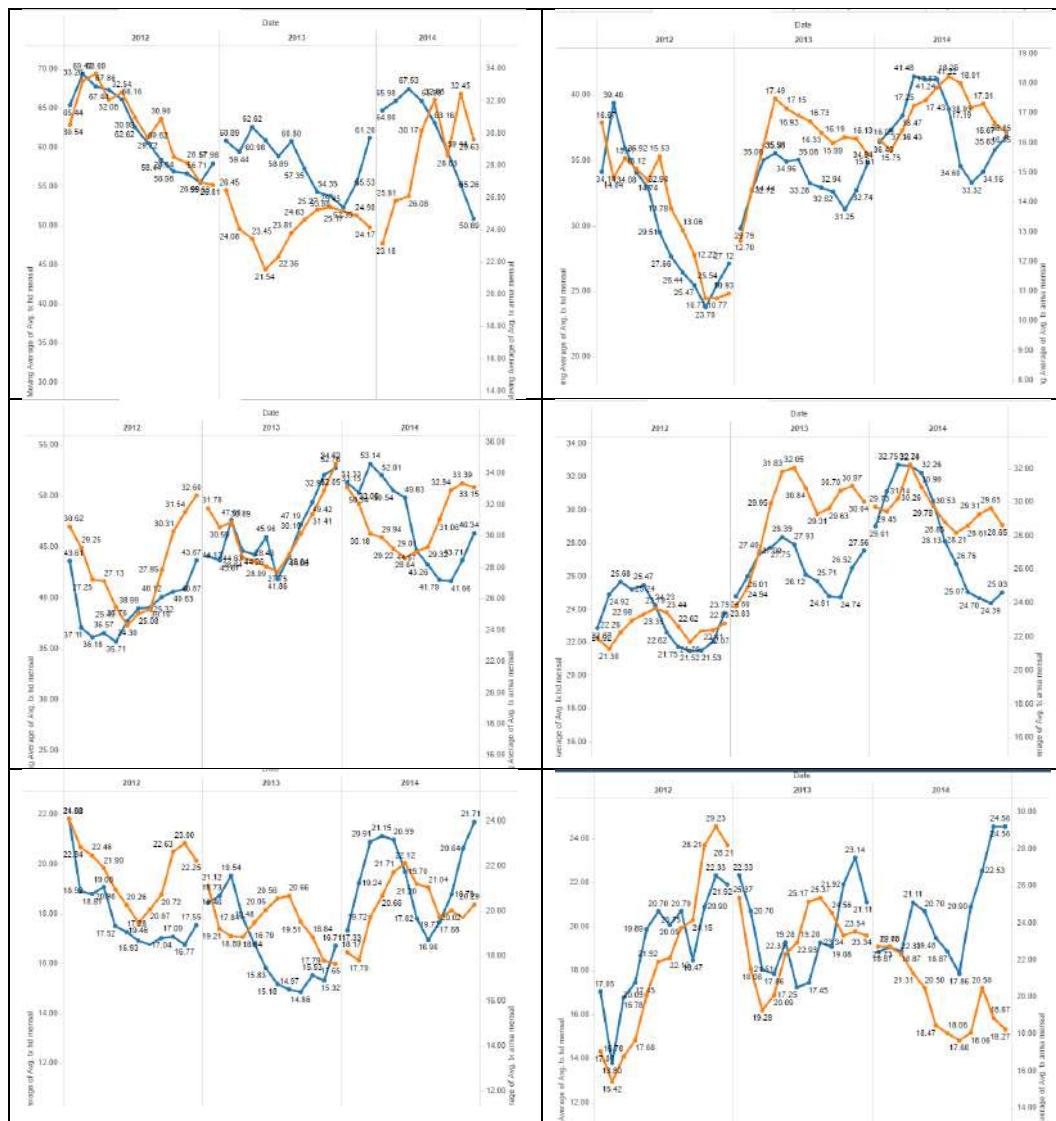
Uma estratégia parecida com a anterior consiste em correlacionar os homicídios (ou a variável de interesse) com outra variável existente na própria base do Sinesp e que se supõe estar associada. Uma das vantagens é que ambas as variáveis estão disponíveis para os mesmos locais e períodos (por ex. por município e por mês). Tome-se, por exemplo, a associação entre homicídios e armas de fogo. Na ausência de uma medida direta da quantidade de armas em circulação, os criminólogos costumam usar *proxys*, ou variáveis substitutas, tais como a porcentagem de suicídios por armas de fogo – PAF, no total de suicídios. Ou ainda as ocorrências de porte ilegal de armas, também coletadas na base Sinesp.

Com relação a esta última medida, alguns questionam se trata-se de uma boa medida substituta para quantidade de armas em circulação, pois ela padece dos problemas de simultaneidade e endogeneidade. Em outras palavras, como as pessoas se armam quando existem mais crimes, tanto homicídios quanto armas em circulação podem aumentar ao mesmo tempo, gerando um falso sinal positivo entre ambas. Além disso, ocorrências de porte

de armas podem estar medindo atividade policial mais intensa, ao invés de mais armas em circulação. Mas como argumentam os adeptos do *Big Data* e do *Data Science*, seguindo a abordagem epidemiológica, pouco importa a causalidade: importa para efeitos simplesmente de previsão que ambas as medidas estejam correlacionadas.

Nos gráficos abaixo vemos a evolução temporal das duas séries de dados do Sinesp – taxa de homicídios dolosos e taxa de porte ilegal de armas – alisadas para um período de 3 meses (média móveis de 3 meses), para 6 Estados: AL, BA, CE, RJ, RS e TO, apenas para ilustrar com casos de todas as regiões.

Gráfico 2 - Evolução temporal das séries de taxa de homicídios dolosos e taxa de porte ilegal de armas do Sinesp para os Estados de AL, BA, CE, RJ, RS e TO. Período de 3 meses.

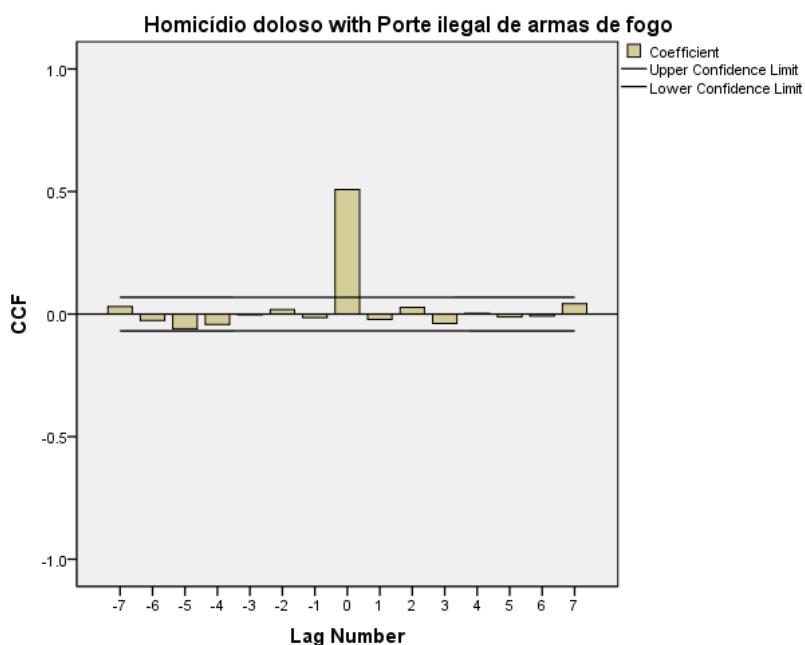


Fonte: Elaboração própria com dados do Sinesp - planilha fornecida pela Coordenação Geral de Pesquisa e Análise da Informação em Segurança Pública - DEPAID/ SENASP/ MJ em 28 de maio de 2015).

Uma análise visual superficial já sugere que os dois fenômenos parecem caminhar aproximadamente juntos no tempo, ainda que com hiatos: regra geral, quando as polícias encontram e apreendem mais armas, temos mais homicídios e vice-versa.

Como existem tendências e sazonalidade nas séries, tomamos as séries em sua diferença ($t-1$) para torná-las estacionárias. O gráfico de correlação cruzada abaixo mostra que a correlação entre as duas séries continua significativa mesmo com as séries diferenciadas ($r^2 = .508$) e que a correlação mais forte se dá no *lag 0*, ou seja, no mesmo mês da apreensão.¹

Gráfico 3 - correlação cruzada entre as séries de taxa de homicídios dolosos e taxa de porte ilegal de armas do Sinesp.

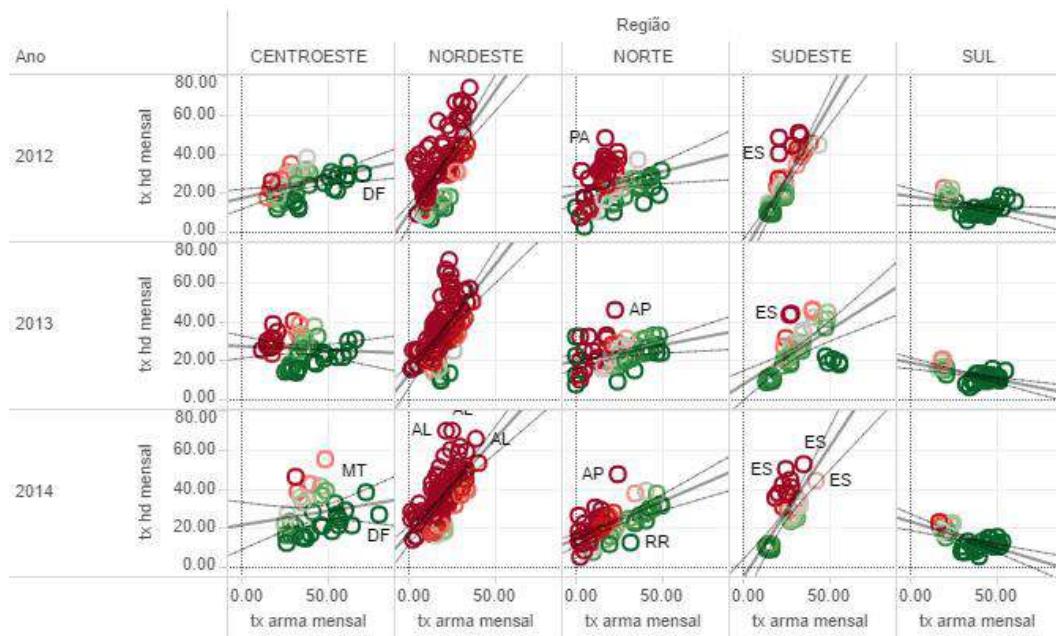


Fonte: Elaboração própria com dados do Sinesp - planilha fornecida pela Coordenação Geral de Pesquisa e Analise da Informação em Segurança Publica - DEPAID/ SENASP/ MJ em 28 de maio de 2015).

Por sua vez, o gráfico de dispersão abaixo, com as taxas de apreensão de armas no eixo horizontal e taxas de homicídios dolosos no eixo vertical – ajuda-nos a ver de modo geral a relação entre as duas variáveis. A relação entre apreensão de armas e homicídios é forte e positiva tanto no Nordeste quanto no Sudeste, nos 3 anos analisados. Ela parece ser um pouco mais fraca, mas ainda positiva no Centro Oeste e no Norte. Finalmente, no Sul, observamos uma associação significativa, mas negativa entre as variáveis.

¹ Uma série temporal é estacionária quando ela se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo alguma forma de estabilidade. A maioria dos procedimentos de análise estatística de séries temporais supõe que as séries sejam estacionárias, portanto, será necessário transformar os dados originais se estes não formam uma série estacionária. A transformação mais comum consiste em tomar diferenças sucessivas da série original, até se obter uma série estacionária.

Gráfico 4 - Gráfico de dispersão para a relação entre apreensão de armas e homicídios, por região.



Fonte: Elaboração própria com dados do Sinesp - planilha fornecida pela Coordenação Geral de Pesquisa e Análise da Informação em Segurança Pública - DEPAID/ SENASP/ MJ em 28 de maio de 2015).

O quadro abaixo traz os coeficientes R quadrado e o nível de probabilidade em cada célula, que corroboram as observações anteriores.

Tabela 8 Coeficientes R para as correlações entre apreensão de armas e homicídios, por região

	CO	NE	N	SE	S
2012	R2 = .18 (p 0.003)	R2 = 0.45 (p. 0.000)	R2 = 0.08 (p. 0.015)	R2= 0.65 (p. 0.000)	R2=0.13 (p.0.024)
2013	R2 = .005 (p. 0.63)	R2 = 0.38 (p. 0.000)	R2 = 0.10 (p. 0.010)	R2= 0.55 (p. 0.000)	R2=0.38 (p. 0.000)
2014	R2 = 0.03 (p. 0.28)	R2 = 0.42 (p. 0.000)	R2 = 0.39 (p. 0.000)	R2= 0.56 (p.0.000)	R2=0.39 (p.0.000)

A associação entre apreensões por porte ilegal de armas e taxa de homicídios é significativa em 12 das 15 células. Na maioria delas, mais armas em circulação implicam em mais crimes, exceto no Sul, onde o sinal é inverso.

Uma conjectura plausível é que quando os crimes patrimoniais aumentam, aumenta também a sensação de insegurança e consequentemente mais pessoas saem armadas nas ruas (quando o risco de ser parado pela polícia é baixo). Com mais armas nas ruas, observamos o crescimento dos homicídios, em especial os de natureza interpessoal. O inverso ocorre quando os crimes patrimoniais diminuem ou quando o risco de ser parado pela polícia é tão elevado que desestimula andar armado. O sinal negativo no Sul é mais complexo e talvez tenhamos que nos valer de estudos qualitativos para compreender a ligação cultural a população da região.

(onde se concentram as fábricas de armas e fronteiras, com histórico de disputas territoriais com os vizinhos) com as armas.

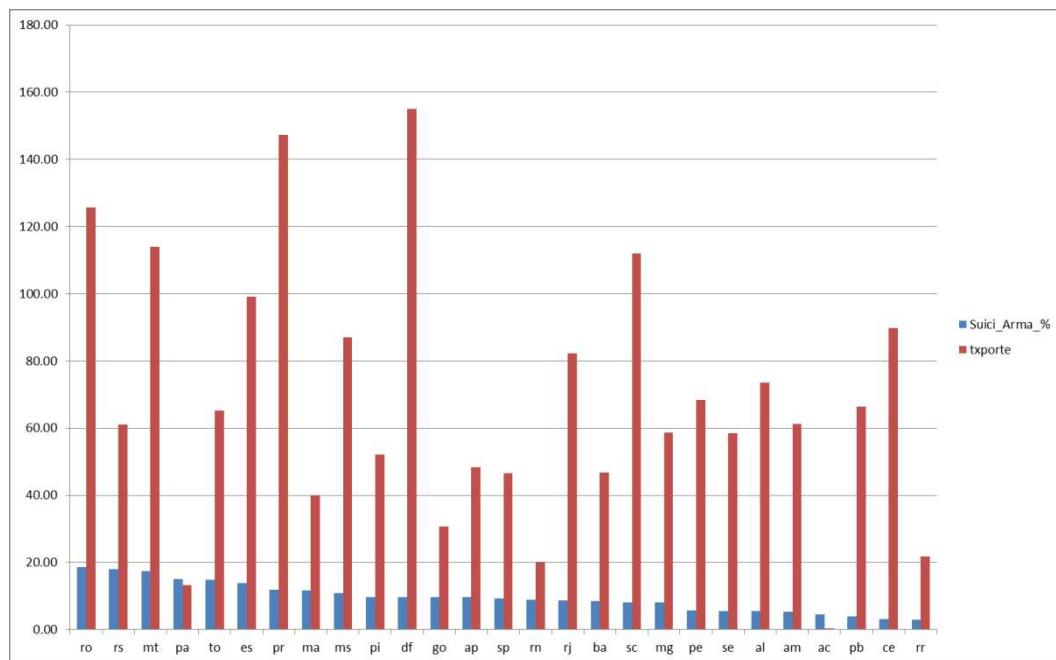
Difícil argumentar que tivemos operações policiais de busca e apreensão de armas em todos estes Estados, nos três anos analisados e que, portanto, as associações observadas devam-se apenas a atividade policial mais intensa. A generalidade da associação por regiões e anos sustenta antes o argumento de que ocorrência de porte ilegal de armas (assim como suicídios com arma de fogo) seja em parte uma medida da quantidade de armas em circulação. O ideal seria calcular uma taxa usando no denominador o número de revistas, de modo a controlar a atividade policial. Este é um exemplo de como é necessário compilar não apenas dados sobre crimes, mas também indicadores de *input*, como revistas, para conseguirmos interpretar corretamente as tendências identificadas.

Não é caso de estender o ponto aqui. A finalidade é apenas mostrar que é possível checar a validade das informações na base de dados valendo-se de variáveis que apresentem elevada correlação estatística. E que é possível extrair informação relevante para traçar políticas públicas mesmo sem ligar a base de crimes com outras variáveis externas, de cunho socioeconômico ou demográfico.

A base Sinesp tem dados inéditos, como a quantidade de apreensão de armas por Estado e por mês, importantes para avaliar uma das principais ações de controle dos homicídios, que passa pelo controle de armas. Até o momento a maior dos estudos que avaliou o impacto do desarmamento sobre a evolução dos homicídios tem se valido da *proxy* de suicídio PAF do SIM/Datasus como substituto da quantidade não observável de armas em circulação. Esta *proxy* já foi validada em estudos anteriores, que mostraram uma correlação, por exemplo, entre a porcentagem de residências com armas em casa e a taxa de suicídio por arma de fogo em diferentes países.

Um modo alternativo de validar a *proxy* porcentagem de suicídio PAF como substituta da quantidade de armas em circulação é ver como ela se correlaciona com a apreensão de armas de fogo pela polícia. No gráfico abaixo vemos a porcentagem de suicídio PAF em 2013 em azul e na coluna vermelha a taxa de apreensão de armas pela polícia por 100 mil habitantes, tomando a média de 2012 a 2014.

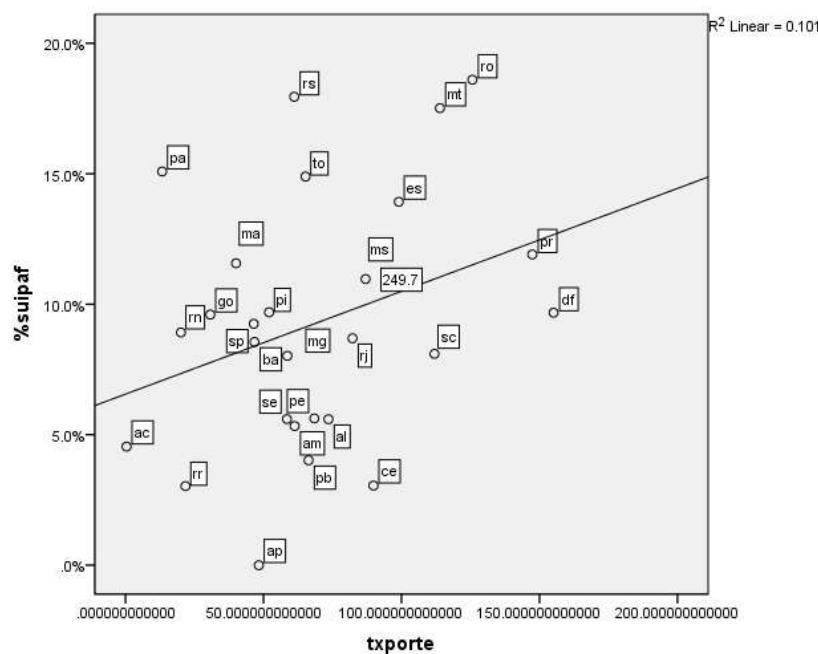
Gráfico 5 - Porcentagem de suicídio PAF (2013) e a taxa de apreensão de armas pela polícia por 100 mil habitantes (média de 2012 a 2014).



Fonte: Elaboração própria a partir dos dados e fontes referidos na tabela 07 e dados do Sinesp - planilha fornecida pela Coordenação Geral de Pesquisa e Analise da Informação em Segurança Publica - DEPAID/ SENASP/ MJ em 28 de maio de 2015).

Como o gráfico sugere, a correlação entre as duas medidas é fraca, atestada também por um baixo coeficiente R =.31. Isto significa que estas variáveis medem coisas um tanto diferentes e que talvez nenhuma delas seja uma *proxy* adequada para número de armas em circulação. É possível que apreensões de armas pela polícia, em alguns Estados, mensurem não apenas a quantidade de armas em circulação, mas também o esforço policial.

Gráfico 6 - Gráfico de dispersão: porcentagem de suicídio PAF (2013) e a taxa de apreensão de armas pela polícia por 100 mil habitantes (média de 2012 a 2014).



Fonte: Elaboração própria a partir dos dados e fontes referidos na tabela 07 e dados do Sinesp - planilha fornecida pela Coordenação Geral de Pesquisa e Analise da Informação em Segurança Publica - DEPAID/ SENASP/ MJ em 28 de maio de 2015).

Trata-se de uma observação preliminar importante uma vez que a porcentagem de suicídio com armas de fogo tem sido utilizada com frequência para estimar o efeito da redução da circulação de armas de fogo sobre os homicídios.

Note pelo diagrama de dispersão acima que em alguns estados, como SP, BA, MG, RJ, SC, DF e PR (casos mais próximos da linha de tendência) – ambas as medidas parecem mensurar coisas semelhantes – provavelmente, quantidade de armas em circulação – enquanto em outros estados há maior discrepância entre os indicadores, talvez porque taxa de apreensão de armas por porte ilegal esteja medindo aí também atividade policial.

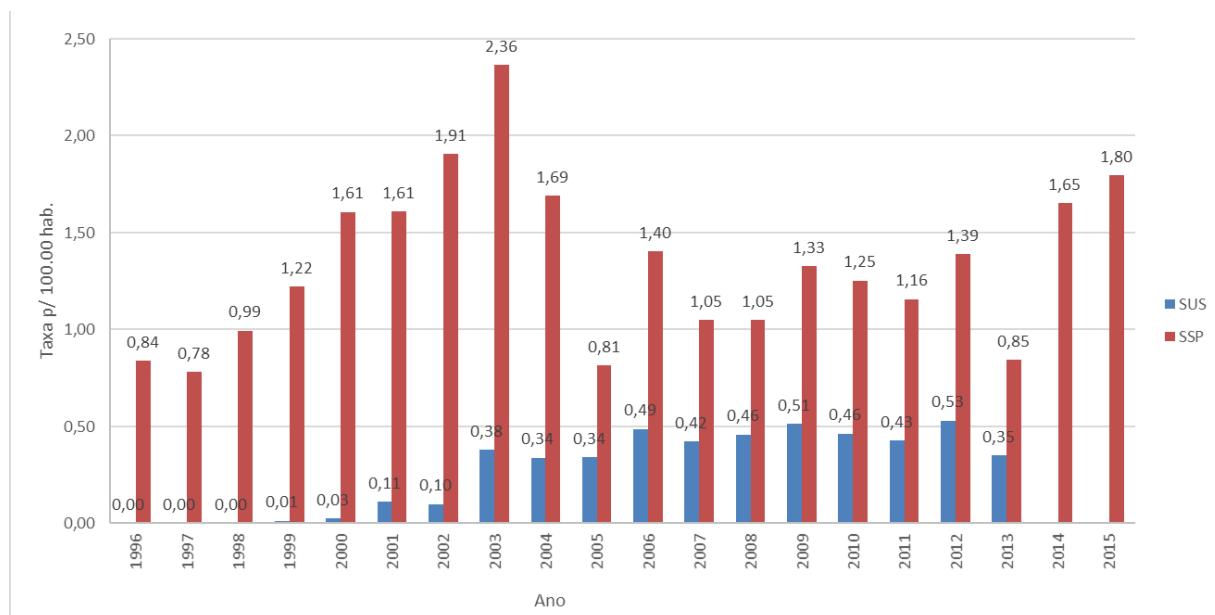
4.3. Letalidade Policial: avaliando políticas públicas

Letalidade policial é uma discussão de grande relevância no Brasil, onde as mortes em confronto com as polícias podem chegar a representar parcelas significativas do total de mortos. Tomando as médias dos anos 2013 e 2014, as mortes em confronto representaram o equivalente a 5% dos homicídios e em Estados como São Paulo, Rio de Janeiro e Santa Catarina podem superar os 10%.² Em São Paulo, para os anos analisados, tivemos cerca de 790 mortes em confronto, mais de 15% das mortes.

Os dados sobre mortes em confronto com as polícias coletados pelo SIM/Datasus são bastante subestimados e a falta de qualidade na coleta praticamente compromete seu uso. Isto torna bastante relevante os dados sobre o fenômeno coletados pelo sistema de justiça criminal, pois são praticamente os únicos que se pode utilizar para monitorar e entender o fenômeno.

No gráfico abaixo, vemos a taxa de letalidade policial entre 1996 e 2015, nas duas fontes, para o Estado de São Paulo. Apesar da melhora na coleta da saúde após 2003, os dados desta fonte refletem em média apenas 24% dos casos registrados pela Secretaria de Segurança Pública.

Gráfico 7 - Taxa de letalidade policial segundo o SUS e a SSP-SP no estado de São Paulo.



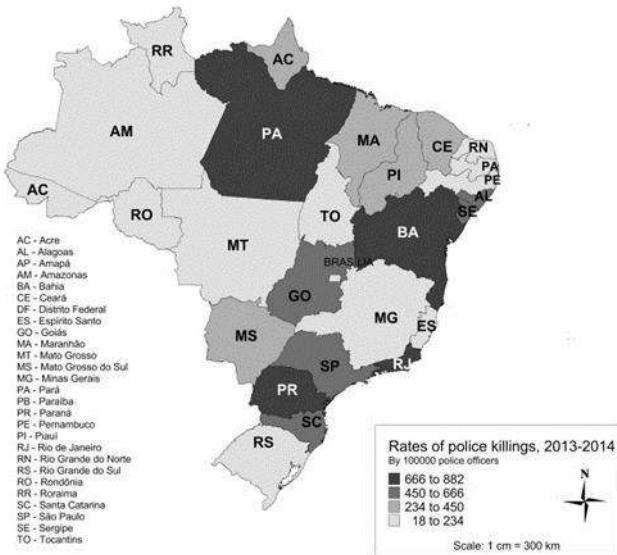
Fonte: Elaboração própria com dados do Datasus: Estatísticas vitais, mortalidade por causas externas, categorias CID-10 Y35 – Y36 (disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937>) e Secretaria de Segurança Pública do Estado de São Paulo.

² A queda dos homicídios em SP e RJ fez aumentar esta porcentagem de mortos em confronto nestes estados, pois tivemos uma queda abrupta no denominador desta porcentagem, nos últimos anos.

Neste tópico procuramos fazer uma análise exploratória das denominadas Mortes em Decorrência de Intervenção Policial (MDIP), conforme definidas pela Senasp e cujos números para 2013 e 2014 podem ser obtidos no Anuário Brasileiro de Segurança Pública³.

O mapa abaixo traz a taxa por 100 mil policiais média de letalidade policial, em 2013-2014, por Estado e mostra taxas elevadas no RJ, PR, PA, BA e SP.

Gráfico 8 - Mapa da taxa média de MDIP em 2013-2014, por Estado.



Fonte: Anuário Brasileiro de Segurança Pública.

Trabalhamos com as médias dos anos de 2013 e 2014, pois como o número absoluto de MDIP é baixo, podemos ter variações bruscas de um ano para outro e as médias ajudam a estabilizar os dados. Na tabela abaixo vemos a média de homicídios dolosos, a média de MDIP e a porcentagem de MDIP no total de mortes em cada Estado. Na coluna seguinte, os Estados estão listados segundo a taxa de letalidade policial por 100 mil policiais. Na última coluna, a taxa de letalidade policial, desta vez calculada com base nas intervenções legais coletadas pelo Datasus.

³ A falta de padronização na coleta de dados sobre MDIP entre as UFs dificulta a apreciação sobre o quanto o número pode ser subestimado. Como nem todas as UFs divulgam o dado de forma sistemática, é razoável supor pelo menos algum grau de subnotificação.

Tabela 9 - Média de homicídios dolosos, a média de MDIP, porcentagem de MDIP no total de mortes e taxa de letalidade policial. Na última coluna, taxa de letalidade policial segundo a categoria intervenção legal do Datasus. Período 2013-2014.

UF	Média de homicídios dolosos	Média de MDIP	Porcentagem de MDIP	Taxa de letalidade policial por 100 mil policiais	Taxa de letalidade policial (Datasus)
Rio de Janeiro	4.742,5	500	10,5%	881,49	414,30
Paraná	2.517	189	7,5%	854,66	108,53
Pará	3.187	155,5	4,9%	831,15	13,36
Bahia	5.530,5	295,5	5,3%	805,20	322,90
São Paulo	4.285,5	789,5	18,4%	648,43	158,92
Sergipe	939	36	3,8%	603,42	8,38
Goiás	2.551	88	3,4%	587,10	13,34
Alagoas	2.098,5	54	2,6%	579,65	5,37
Santa Catarina	730	73,5	10,1%	498,27	189,82
Mato Grosso do Sul	526	32	6,1%	444,94	20,86
Maranhão	1.734,5	41	2,4%	420,81	30,79
Amapá	229	14,5	6,3%	301,96	10,41
Ceará	4.237	47	1,1%	254,03	5,40
Piauí	574	16,5	2,9%	240,17	29,11
Espírito Santo	1.542	23	1,5%	209,55	18,22
Rio Grande do Sul	2.071,5	53,5	2,6%	206,21	42,40
Minas Gerais	3.924	91,5	2,3%	176,44	11,57
Amazonas	915	18	2,0%	159,11	4,42
Paraíba	1.467	17,5	1,2%	158,16	4,52
Rondônia	480,5	12	2,5%	157,34	19,67
Pernambuco	3.077,5	36,5	1,2%	143,91	1,97
Tocantins	308	6	1,9%	108,42	18,07
Mato Grosso	1.133	8	0,7%	89,24	33,46
Acre	190	2	1,1%	52,66	13,16
Roraima	85,5	1	1,2%	39,56	19,78
Distrito Federal	686	4,5	0,7%	23,77	10,56
Rio Grande do Norte	1.435	1	0,1%	18,42	4,61
	51196,5	2606,5	5,1%		

Fonte: Elaboração própria com dados do Datasus: Estatísticas vitais, mortalidade por causas externas, categorias CID-10 Y35 – Y36 (disponível em <http://www2.datasus.gov.br/DATASUS/index.php?area=0205&id=6937>) e Anuário Brasileiro de Segurança Pública.

Como pode ser visto na tabela, São Paulo e Rio de Janeiro têm as maiores porcentagens de mortes em confronto no total de mortes. Os índices de letalidade policial são tradicionalmente elevados nestas UFs, mas note-se que estes são precisamente os Estados com maiores quedas nos homicídios dolosos. Assim, parte da responsabilidade pela posição proeminente na lista se

deve à diminuição no denominador (homicídios dolosos) e não necessariamente a aumentos no nominador (mortes em confronto). Nesse caso, significa dizer que o padrão de letalidade policial era alto e continua alto, destacando-se ainda mais a partir da redução das demais mortes violentas intencionais. Outra hipótese explicativa é de caráter metodológico, ou seja, SP e RJ publicam tais dados faz algumas décadas, enquanto outras Unidades da Federação ainda os tratam como informação reservada. Nesse processo, os dados do RJ e de SP apresentariam maior estabilidade e série histórica.

A tabela mostra também como as taxas de letalidade por 100 mil policiais calculadas com base nos dados da segurança são bem superiores as taxas calculadas com o Datasus. Ainda assim, as duas séries mostram alguma congruência, não obstante as grandes discrepâncias como os casos do Pará, Sergipe ou Alagoas. Como discutido anteriormente, os dados da saúde para a análise deste fenômeno são pouco confiáveis e mostramos aqui apenas como ilustração do problema.

Neste estudo exploratório, utilizamos como variável independente a taxa de mortes em confronto por 100 mil policiais para controlar o efeito do tamanho das polícias nos Estados e uma série de variáveis independentes para tentar compreender as variações observadas entre os Estados. A tabela abaixo descreve o rol de variáveis preditoras utilizadas no modelo e suas fontes.

Tabela 10 - Variáveis preditoras utilizadas no modelo e suas fontes.

Variável	Observação	Fonte
UF	Código UF	IBGE
Letalidade_Policial_TXP	(Letalidade policial 2013+2014)/2)/n. de policiais * 100000 - Variável	FBSP
Policiais_Mortos_TXP	(Policiais mortos 2013+2014)/2)/n. de policiais * 100000	FBSP e IBGE
Roubo_Carga	(Roubos 2013+2014)/2)/n. de hab * 100000	SINESP - FBSP
Confianca_PM_%	Pesquisa de vitimização (2013)	SENASP, CRISP e DataFolha
Custo_SP	Custo em Segurança Pública per capita	FBSP
TXAVGHD	TAXA DE HOMICÍDIOS COM BASE NA MÉDIA 2013 E 2014	SINESP - FBSP
Percepcao_Crime_%	Quanto o crime aumentou em %. Proxy para medo/insegurança	Vitimização Senasp
Suici_Arma_%	% de suicídios por arma de fogo. Proxy para armas	DATASUS
TX_PRHD	TAXA DE PRISÃO DE HOMICIDAS	DEPEN
HDI	Índice	PNUD (2010)
Corregedoria	Possui corregedoria - Dummy	Estadic/IBGE (2014)
Ouvidoria	Possui ouvidoria - Dummy	Estadic/IBGE (2014)
Recurso_FNSP	Recurso do Fundo Nacional de Segurança Pública - Dummy	Estadic/IBGE (2014)
PRPH	Plano de redução e/ou prevenção de homicídios - Dummy	Estadic/IBGE (2014)
Cap_Uso_Forca	Capacitação Uso diferenciado da força - Dummy	Estadic/IBGE (2014)
Cap_Direitos_Hum	Capacitação em Direitos Humanos - Dummy	Estadic/IBGE (2014)

Nos confrontos entre policiais e civis morrem geralmente mais civis, uma vez que os policiais são mais bem preparados e equipados que os criminosos para tais situações. Ainda assim, como se trata de um confronto entre dois lados, espera-se naturalmente que exista uma correlação positiva entre as duas grandezas: onde a polícia mata mais, provavelmente, também morre mais. Entretanto, conforme vários estudos do Fórum Brasileiro de Segurança Pública já apontaram, a quantidade de mortes é diferente do risco relativo. Se tomarmos as quantidades de mortes de policiais e de não policiais segundo o tamanho dos efetivos e da população, os policiais correm em média três vezes mais chances de serem mortos do que os

não policiais. Esse fenômeno reforça a ideia de que as mortes violentas intencionais são um grave problema que não atinge apenas um segmento da sociedade brasileira mas a todos.

Queríamos uma variável que mensurasse a importância do crime organizado no Estado e na falta de uma variável diretamente observável usamos roubo de carga como uma aproximação. Em tese ela é melhor do que tráfico de drogas, pois embora esta seja também uma atividade típica de crime organizado, o indicador mede antes a atividade policial do que o fenômeno retratado, pois tráfico é um tipo de crime cometido contra a sociedade, mas sem vítimas no sentido estrito. Com exceção dos períodos de crise – como os ataques de 2006 e 2012 em São Paulo – a literatura sugere que quanto mais organizado o crime, menos violento. O crime organizado típico (como jogo do bicho) prefere corromper a polícia a trocar tiros com ela. Deste modo, esperamos um sinal negativo entre crime organizado e letalidade policial.

Confiança na PM e Percepção do crime (“crime está aumentando na cidade nos últimos 5 anos”) são questões oriundas da pesquisa nacional de vitimização da Senasp, agregadas estadualmente. O esperado é que uma polícia pouco letal e respeitadora dos direitos fundamentais goze de mais confiança da população de modo que deveríamos encontrar uma associação inversa entre letalidade e confiança na PM. Por outro lado, como parte da população apoia explicitamente os abusos quando cometidos contra “bandidos”, não seria espantoso se o sinal fosse invertido. Do mesmo modo, a percepção de que o crime está aumentando pode ser menor nos estados com elevada letalidade, pois a notoriedade dos casos daria a sensação de que a polícia é enérgica contra os bandidos.

Gastos em segurança: a conjectura aqui é de que a violência policial (tortura, execuções, etc.) é um recurso que polícias pouco capacitadas e equipadas utilizam para dar uma resposta à sociedade frente às cobranças por segurança. Assim, quanto mais recursos o Estado investe na polícia, mais preparada e técnica deveria ser sua atuação, implicando em menores taxas de letalidade.

Taxa de homicídio e porcentagem de suicídio com arma: a suposição por traz da escolha da taxa de homicídio é que nos estados com muitos homicídios os policiais também temem a violência. O medo do policial pode fazer com que ele reaja antecipadamente, escolhendo por “matar antes de morrer”. Isto explicaria em parte por que os confrontos tendem a ocorrer nos bairros que concentram as maiores taxas de homicídio. Por outro lado, esta relação talvez seja mais tênue nos locais onde os homicídios têm natureza mais interpessoal e menos ligados à dinâmica criminal. Porcentagem de suicídios cometidos com arma de fogo é na verdade uma variável substituta para quantidade de armas em circulação, que está diretamente ligada à quantidade de homicídios. A hipótese aqui é que quanto mais armas (e homicídios), maior a quantidade de confrontos com a polícia e, portanto, maior o número de vítimas fatais.

Taxa de prisão de homicidas: trata-se de uma medida de “impunidade”. Entrevistas com policiais que se envolveram em confrontos dão conta da queixa generalizada entre os policiais militares de que eles prendem e a polícia civil ou a justiça soltam os criminosos. De que as leis são pouco rigorosas e que os criminosos voltam rapidamente para as ruas, por falta de investigação eficiente, corrupção ou artifícios legais. Muitos se viram tentados, assim, a “fazer

justiça com as próprias mãos”, forçando os confrontos quando eventualmente outro desfecho seria possível (MANSO, 2012).

A hipótese aqui é que nos estados com uma menor impunidade com relação aos homicidas, são menores os incentivos ao confronto, pois haveria maior credibilidade quanto ao funcionamento do sistema de justiça criminal. O ideal seria termos uma medida de impunidade policial (ex.: policiais presos após mortes ilegais em confronto), mas esta medida não está disponível.

Índice de Desenvolvimento Humano: o IDH é uma medida sintética de “desenvolvimento”, abarcando conjuntamente as dimensões expectativa de vida, escolaridade e renda. Sua introdução na equação visa a controlar as disparidades de desenvolvimento social e econômico entre os Estado. Mas qual o sinal esperado? IDH contem renda em sua composição e o efeito da renda sobre a criminalidade é dubio: em geral, os crimes patrimoniais oportunistas aumentam com o aumento da renda enquanto os crimes homicídios caem. Não temos neste caso hipóteses sobre o sentido da associação com a letalidade policial, se é que ela existe.

Corregedoria e Ouvidoria: são mecanismos institucionais clássicos de controle interno e externo das atividades policiais. O esperado é que policias estaduais com estes órgãos tenham índices menores de letalidade, na medida em que a punição aos eventuais abusos atue como inibidor para o envolvimento em confrontos.

Recursos do Fundo Nacional de Segurança e Plano de Redução de Homicídios: os recursos do fundo nacional de segurança pública podem ser um instrumento para o direcionamento da política nacional de segurança pública, que nas últimas décadas tem insistido com os estados para a necessidade de redução da letalidade policial. Esperaríamos assim menores taxas de letalidade dos estados que recebem recursos do FNSP, bem como daqueles que elaboraram algum Plano de redução de homicídios.

Capacitação: Os Estados que capacitam seus policiais com cursos de uso diferenciado da força ou direitos humanos deveriam, em tese, ter menores taxas de letalidade policial. Não se sabe pela Estadic quando, de que forma e com que intensidade os cursos foram adotados e se tiveram algum efeito, mas deveríamos esperar encontrar uma associação negativa com a variável dependente.

Todas estas variáveis foram coletadas para os 27 UFs num *design cross-sectional* (comparação entre os Estados, num momento do tempo) e posteriormente incluídas numa regressão OLS (ou MMQ, mínimos quadrados ordinários) usando a estratégia *enter*, ou seja, todas deveriam estar forçosamente presentes no modelo, mesmo que o nível de significância sugerisse sua exclusão por ausência de peso no resultado final. Não tínhamos *a priori* interesse em selecionar apenas as variáveis com maior poder de explicação, pois acreditamos que numa análise exploratória, a ausência de peso para algumas variáveis que acreditamos relacionadas à letalidade é também um resultado substantivamente relevante. Nossa preocupação não era fazer uma previsão acurada, mas antes analisar a contribuição de cada variável para o modelo.

Tabela 11 - Resumo do modelo**Model Summary^b**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics						Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change		
1	.913 ^a	.833	.638	168.8191891	.833	4.277	14	12	.008	2.426	

a. Predictors: (Constant), Cap_Direitos_Hum, PRPH, Percepcao_Crime_%, TXAVGHD, Custo_SP, TX_PRHD, Policias_Mortos_TXP, Ouvidoria, Recurso_FNSP, Suici_Arma_%, Confianca_PM_%, HDI, Corregedoria, Cap_Uso_Forca

b. Dependent Variable: Letalidade_Policial_TXP

Tabela 12 - Teste ANOVA**ANOVA^b**

Model	Sum of Squares		df	Mean Square	F	Sig.
	Regression	Residual				
1	1706511.420	341999.023	14	121893.673	4.277	.008 ^a
	Total	2048510.443	26	28499.919		

a. Predictors: (Constant), Cap_Direitos_Hum, PRPH, Percepcao_Crime_%, TXAVGHD, Custo_SP, TX_PRHD, Policias_Mortos_TXP, Ouvidoria, Recurso_FNSP, Suici_Arma_%, Confianca_PM_%, HDI, Corregedoria, Cap_Uso_Forca

b. Dependent Variable: Letalidade_Policial_TXP

Tabela 13 - Coeficientes**Coefficients^a**

Model	Unstandardized Coefficients		Beta	t	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
	B	Std. Error			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1 (Constant)	-1816.590	998.870		-1.819	.094	-3992.940	359.761				
Policias_Mortos_TXP	2.741	.971	.512	2.821	.015	.624	4.857	.320	.632	.333	.423
Confianca_PM_%	16.215	14.914	.233	1.087	.298	-16.280	48.710	-.136	.299	.128	.302
Percepcao_Crime_%	3.301	504.825	.001	.007	.995	-1096.618	1103.219	-.025	.002	.001	.522
Custo_SP	-.697	.318	-.295	-2.194	.049	-1.389	-.005	-.163	-.535	-.259	1.298
TXAVGHD	7.475	3.637	.323	2.055	.062	-.449	15.398	.136	.510	.242	1.779
Suici_Arma_%	-25.167	10.434	-.400	-2.412	.033	-47.900	-2.434	.040	-.571	-.285	.506
TX_PRHD	-6.795	1.394	-.721	-4.875	.000	-9.832	-3.758	-.501	-.815	-.575	.636
HDI	2522.581	1116.811	.443	2.259	.043	89.258	4955.904	.023	.546	.266	.362
Corregedoria	-15.915	126.530	-.026	-.126	.902	-291.601	259.771	-.161	-.036	-.015	.316
Ouvidoria	-37.007	112.437	-.056	-.329	.748	-281.987	207.972	-.022	-.095	-.039	.483
Recurso_FNSP	316.713	119.096	.408	2.659	.021	57.226	576.201	.239	.609	.314	.590
PRPH	241.059	113.302	.400	2.128	.055	-5.805	487.924	.372	.523	.251	.394
Cap_Uso_Forca	212.669	153.637	.364	1.384	.191	-122.077	547.415	.106	.371	.163	.201
Cap_Direitos_Hum	-149.180	158.724	-.247	-.940	.366	-495.011	196.650	.147	-.262	-.111	.201

a. Dependent Variable: Letalidade_Policial_TXP

Fonte: Elaboração própria a partir dos dados e fontes referidos na tabela 10.

O resumo do modelo mostra o R2 elevado (.83) mas este também é um reflexo da inclusão forçada das 14 variáveis preditoras, que terminam por explicar em conjunto a maior parte da variância encontrada e inflacionar o coeficiente. O modelo é por assim dizer saturado e este R2 cai de magnitude se retiramos os fatores estatisticamente pouco significativos, como "Confiança na PM", "Percepção_crime%", existência de "Corregedoria" ou "Ouvidoria", Capacitação em Uso da Força ou em Direitos Humanos. Mas, como argumentamos, para nós o mais interessante aqui é sugerir que estas variáveis parecem simplesmente influenciar muito pouco o nível de letalidade policial nos Estados. Dito de outra forma, as mortes provocadas por intervenção policial parecem não ser fortemente afetadas pelas ações institucionais que têm sido implementadas nos últimos anos no Brasil.

O modelo também apresenta um índice de colinearidade (Durbin-Whatson) elevado o que sugere que algumas variáveis presentes são redundantes (multicolinearidade): talvez melhorássemos os parâmetros do modelo ao escolher entre homicídios ou armas, capacitação em uso da força ou direitos humanos, corregedoria ou ouvidoria, etc., mas como dissemos a estratégia aqui é exploratória e nesta primeira abordagem é relevante saber como cada um dos fatores se comporta na presença dos demais. Mesmo sofrendo de um tanto de saturação e multicolinearidade, o modelo geral é bastante razoável e nos mostra varias coisas interessantes:

- Como esperado, quanto mais mortos civis nos confrontos, mais policiais mortos. A relação é positiva e estatisticamente significativa. O achado reforça a hipótese da ameaça, segundo a qual a letalidade policial é maior em contextos em que o policial se sente ameaçado.
- As duas variáveis proveniente da pesquisa nacional de vitimização e de natureza subjetiva, Confiança na PM e Percepção do crime, aparentemente não influenciam a distribuição da taxa de letalidade entre os estados. Não podemos afirmar, portanto que uma sensação de insegurança generalizada leve os policiais a quererem “mostrar serviço” através da morte de criminosos. Neste sentido, é salutar que a corporação seja “neutra” com relação a estas pressões populares por vingança, não obstante vários discursos de ódio que têm ganhado destaque no país mais recentemente;
- O gasto em segurança per capita (Custo_SP) mostrou-se como postulado a princípio significativo e com sinal negativo. Em outras palavras, quanto maiores os investimentos do estado em segurança (o que em tese implica numa polícia mais técnica, treinada e equipada) menores as taxas de letalidade policial. Do ponto de vista das políticas públicas, é boa a sugestão de que um incremento nos gastos contribuiria em tese para a redução da letalidade;
- Como esperado, a letalidade policial é maior nos estados com maiores taxas de homicídio. O contexto violento, como arguido, pode aumentar o temor dos policiais e com isso estimular o confronto, quando outras medidas seriam cabíveis. Uma alternativa é que o número elevado de mortes provoque uma dessensibilização geral com relação às mortes, de todos os tipos, tornando a comunidade mais complacente com as mortes em confronto. Curiosamente, nossa proxy para quantidade de armas em circulação, revelou-se significativa, mas negativa, sinal contrário ao esperado. São Paulo e Rio de Janeiro tiveram queda na quantidade de armas em circulação (medida aqui pela proxy % de suicídios com arma) mas ainda são estados no topo da letalidade;
- A medida de impunidade foi a mais significativa do modelo, implicando que quanto maior a taxa de prisão de homicidas, menor é a letalidade policial. A punição dos homicidas aventa-se aqui, levaria os policiais acreditarem no funcionamento do sistema de justiça criminal e reduziria em contrapartida a disposição a fazer justiça com as próprias mãos. Vemos aqui outra política pública capaz de gerar algum impacto na redução da letalidade;
- O índice de desenvolvimento mostrou-se significativo e positivo, indicando que quanto mais desenvolvido o Estado, maior a taxa de letalidade policial. O resultado pode parecer surpreendente, mas como mencionamos o IDH tem renda em sua composição. Assim, se

crimes patrimoniais oportunistas crescem com a renda e a melhoria da qualidade de vida da sociedade e se os confrontos ocorrem em sua maioria em incidentes envolvendo roubos em andamento, faz sentido esperarmos um sinal positivo entre IDH e renda: mais renda = mais roubos = mais confrontos = mais letalidade. Esse é um padrão que muitas secretarias reconhecem e indica a necessidade de se discutir como lidar com roubos sem o recurso do enfrentamento aberto e, consequentemente, de maior risco de mortes por intervenção policial;

- Controle institucional: o sinal está negativo, como esperado: a existência de corregedoria e ouvidoria no estado tende a diminuir a taxa de letalidade. Todavia, mesmo a interpretação do sentido da relação é arriscada uma vez que ambas mostraram-se muito pouco significativas no modelo. No que tange à capacitação, o mesmo ocorre com a “capacitação e direitos humanos”. O sinal é negativo como postulado, mas não significativo. Nem as instituições de controle nem a capacitação em temas correlatos parecem ajudar a explicar a atual distribuição da letalidade nos estados. Não estamos querendo dizer que elas não funcionem. Pensando contra factualmente, os níveis de letalidade poderiam ser maiores na ausência destes mecanismos. Eles simplesmente não nos ajudam a entender porque alguns estados têm taxa de letalidade maior do que outros.

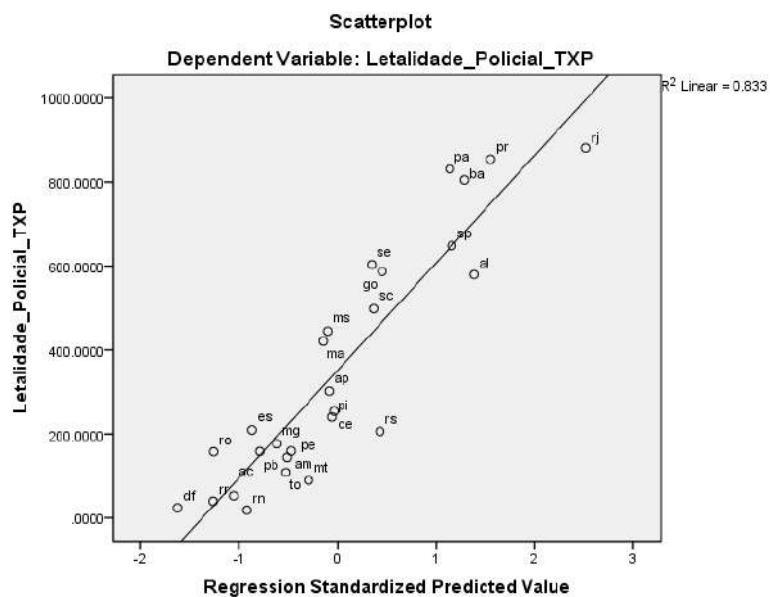
- Recursos do Fundo Nacional de Segurança Pública e Plano de Redução de Homicídios. Estas duas variáveis revelaram-se significativas no modelo, mas com sinal positivo, ou seja, a letalidade é maior nos Estados que declararam receber recursos do FNSP e que contam com algum plano para a redução dos homicídios. Aparentemente, Estados com letalidade elevada não vêm sendo punidos com restrições ao acesso ao FNSP, mas antes, premiados. Por outro lado, talvez sejam estes precisamente os estados que mais necessitam de recursos para capacitar e equipar melhor suas polícias, para que num futuro próximo a taxa de letalidade caia. O problema é que a capacitação, da forma como é feita hoje, não muda práticas institucionais e mostra pouca correlação com o fenômeno. A hipótese é de que se faz necessário ajustar as regras do FNSP e o conteúdo das capacitações por ele apoiadas.

Vimos acima que existe uma relação significativa e positiva entre taxa de homicídios e taxa de letalidade. Faz sentido, portanto, que os Estados que têm as maiores taxas de homicídio sejam aqueles que também tenham planos para a redução dos homicídios, resultando daí a associação positiva entre existência de plano e letalidade.

Somente um estudo de tipo painel poderia avaliar a eficácia desta e de outras políticas para a redução da letalidade. Num modelo *cross-sectional* como o rodado podemos apenas verificar que neste momento específico, são os Estado que recebem recursos e tem plano de redução de homicídio precisamente os que têm maiores taxas de letalidade policial.

O gráfico de dispersão abaixo mostra São Paulo exatamente na posição prevista pela equação. É interessante esta observação, pois mesmo que encabeçando a lista dos estados com policiais mais letais, ele se comporta como esperado, em função de suas características e o Rio de Janeiro, líder em letalidade, aparece como ligeiramente abaixo do esperado. Por outro lado, vemos diversos estados com taxas baixas de letalidade, mas ainda assim acima do esperado, em função das características analisadas.

Gráfico 9 - Gráfico de dispersão: Letalidade policial observada X Letalidade predita pelo modelo



Fonte: Elaboração própria a partir dos dados e fontes referidos na tabela 10.

O modelo nos ajuda a perceber de outra maneira quais os estados “normais” e quais os anômalos, quando controlamos simultaneamente diversos fatores. Na lista abaixo é possível conferir a taxa de letalidade policial observada, a taxa previda e o resíduo estandardizado, que é a diferença entre ambas, colocadas em escore Z. A tabela está organizada por ordem de resíduo, de modo que os Estados no início da tabela têm taxas de letalidade maiores do que as esperadas e os Estados no final têm taxas menores do que as esperadas, segundo o modelo.

Tabela 14 - Taxa de letalidade policial

UF	Letalidade_Policial_TXP	Unstandardized Predicted Value	Standardized Residual
PA	831.1508	643.29699	1.11275
SE	603.4194	440.17260	.96699
RO	157.3358	29.51254	.75716
BA	805.1991	681.57481	.73229
GO	587.0972	467.32370	.70948
MS	444.9388	326.05698	.70420
MA	420.8149	314.56843	.62935
PR	854.6622	749.67211	.62191
DF	23.7705	-64.27237	.52152
ES	209.5481	128.75142	.47860
SC	498.2713	444.85133	.31643
RR	39.5570	28.27472	.06683

PB	158.1563	149.02266	.05410
SP	648.4280	648.17466	.00150
MG	176.4400	194.23555	-.10541
AP	301.9575	330.02642	-.16627
AC	52.6593	80.92494	-.16743
AM	159.1090	230.32455	-.42185
PE	143.9104	220.57104	-.45410
CE	254.0266	342.24028	-.52253
PI	240.1747	336.47613	-.57044
RN	18.4247	116.00930	-.57804
TO	108.4207	217.26151	-.64472
RJ	881.4922	995.72222	-.67664
AL	579.6479	705.64053	-.74632
MT	89.2359	276.16452	-1.10727
RS	206.2054	461.47613	-1.51210

Fonte: Elaboração própria a partir dos dados e fontes referidos na tabela 10.

Em linhas gerais, o exercício proposto sugere que a letalidade é influenciada pelo nível geral de mortes violentas intencionais no Estado. Vimos ainda que ela responde positivamente a incentivos como aumento dos investimentos em segurança ou diminuição da impunidade pelo aprisionamento de homicidas, mas que talvez o controle institucional e a capacitação estejam falhando no propósito de controlar e aprimorar a atividade policial. Por razões que precisariam ser aprofundadas, o recebimento de recursos do Fundo de Segurança e a criação de planos de redução de homicídios estão positivamente relacionados à letalidade policial, o que não implica de maneira alguma em causalidade.

Vamos parar a análise neste ponto uma vez que a finalidade destes três ensaios não era esgotar os temas propostos, sobre os quais muito já se escreveu e que exigiriam muitos outros dados e procedimentos. Como explicitado inicialmente, a proposta era antes ilustrar de que modo a base de dados do Sinesp poderia ser utilizada pelos gestores públicos e pela comunidade acadêmica para gerar insights sobre a relevante questão dos homicídios no Brasil.

5. CONSIDERAÇÕES FINAIS

A qualidade de uma base de dados pode ser verificada através de diferentes metodologias. Uma alternativa sugerida aqui é a identificação de variáveis tradicionalmente associadas aos homicídios para verificar a tanto a relação geral quanto os casos extremos, através da comparação dos números observados com os números previstos pela combinação linear de uma série de variáveis preditoras. Chamamos este procedimento de validação externa e mostramos que a taxa de homicídio doloso coletada pelo Sinesp pode ser razoavelmente previda através de uma combinação de variáveis socioeconômicas e demográficas do IBGE. Em outras palavras, a distribuição dos homicídios pelo território não é aleatória, mas guarda relação com variáveis já identificadas em outros estudos, o que de algum modo corrobora a fidedignidade da base do Sinesp.

Um modo alternativo é correlacionar os homicídios com outra variável pertencente à própria base de dados – como porte ilegal de armas – pois ainda que mensurem fenômenos diversos, é comum que ambas caminhem juntas. Novamente aqui, constatamos uma associação significativa entre as taxas de homicídio nos Estados e as taxas de apreensão de armas pela polícia, corroborando a literatura que identifica a quantidade de armas em circulação como um dos principais preditores dos homicídios. Vimos, por outro lado, que quantidade de armas apreendidas não se correlaciona fortemente com outra variável tradicionalmente usada para medir armas em circulação, os suicídios PAF. Ao contrário de suicídios, é possível que a variável apreensão de armas seja bastante endógena, medindo tanto armas em circulação quanto atividade policial.

O exercício sobre letalidade policial proporciona uma série de *insights* sofre fatores potencialmente associados aos maiores ou menores graus de violência policial nos diferentes Estados. Mais do que por suas conclusões substantivas, o exercício ilustra uma situação em que os dados coletados pelo sistema de justiça criminal são bastante superiores aos dados provenientes do sistema de saúde, cujas informações sobre mortes decorrentes de intervenções legais são bastante subestimadas.

De modo geral, ao menos no que tange aos homicídios, a base de dados do Sinesp parece bastante consistente, pois como evidenciamos, está bastante associada à base de agressões externas proveniente do sistema de saúde, está associada a uma série de variáveis externas preditoras de homicídios bem como à variável interna “porte ilegal de armas”.

O relatório mostra como uma base nacional de indicadores criminais e de atividade policial é fundamental para o estudo da criminalidade e avaliação de políticas públicas. Dados como apreensão de armas pela polícia em nível estadual e mensal são inéditos e aqui pela primeira vez mostramos sua relação com os homicídios. Dados sobre mortes por intervenção policial são de qualidade muito superior aos produzidos pelo Datasus. Isto só reforça a relevância social de disponibilizar de modo periódico e sistemático este conjunto de informações para a sociedade, que agrupa valor ao analisá-los.

6. REFERÊNCIA BIBLIOGRÁFICA

- ANDRADE, Selma Maffei de, et al. "Homicídios de homens de quinze a 29 anos e fatores relacionados no estado do Paraná, de 2002 a 2004." *Cien Saude Colet* 16.Supl 1 (2011): 1281-1288.
- BARATA, Rita Barradas, Manoel Carlos Sampaio de Almeida Ribeiro, and Meri De Sordi. "Desigualdades sociais e homicídios na cidade de São Paulo, 1998." *Revista Brasileira de Epidemiologia* 11.1 (2008): 3-13.
- CERQUEIRA, Daniel. Mortes violentas não esclarecidas e impunidade no Rio de Janeiro. *Econ. Apl.* [online]. 2012, vol.16, n.2 [cited 2016-07-05], pp.201-235
- CORREIA, Fabio Leandro dos Santos. "Homicídio e uso de álcool: relação com as desigualdades sociais." (2014).
- DOS SANTOS, Marcelo Justus, and Ana Lúcia Kassouf. "Estudos econômicos das causas da criminalidade no Brasil: Evidências e controvérsias." *Revista Economia* 9.2 (2008): 343-372.
- DUARTE, Elisabeth Carmen, et al. "Ecological association between characteristics of the municipalities and the risk of homicide in adult males aged 20 to 39 in Brazil: 1999-2010." *Ciência & Saúde Coletiva* 17.9 (2012): 2259-2268.
- MANSANO, Nereu Henrique, et al. "Homicídios em homens jovens de 10 a 24 anos e condições sociais em municípios do Paraná e Santa Catarina, Brasil, 2001-2010." *Epidemiologia e Serviços de Saúde* 22.2 (2013): 203-214.
- MANZO, Bruno Paes. Crescimento e queda dos homicídios em SP entre 1960 e 2010. Uma análise dos mecanismos da escolha homicida e das carreiras no crime [online]. São Paulo : Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, 2012. Tese de Doutorado em Ciência Política.
- MINISTÉRIO DA SAÚDE, SECRETARIA DE VIGILÂNCIA EM SAÚDE DEPARTAMENTO DE ANÁLISE DA SITUAÇÃO DE SAÚDE. Manual para investigação do óbito com causa mal definida, 2008.
- NERY, Marcelo Batista, et al. "Indicadores do risco: análise dos homicídios por intermédio de variáveis preditoras no município de São Paulo em 2001." *Simpósio Brasileiro de Sensoriamento Remoto* 13 (2005).
- PERES, Maria Fernanda Tourinho, et al. "Queda dos homicídios no município de São Paulo: uma análise exploratória de possíveis condicionantes." *Rev bras epidemiol* 14.4 (2011): 709-721.
- SOARES, Gláucio Ary Dillon, Eduardo Cerqueira BATITUCCI, and Ludmila Mendonça Lopes Ribeiro. "As mortes desiguais em Minas Gerais: gênero, idade, estado civil e raça nos homicídios registrados pelo Sistema de Informações de Mortalidade." *Homicídios no Brasil* (2007).
- SOUZA, Carlos Augusto Moreira de, Cosme Marcelo Furtado Passos da Silva, and Edinilda Ramos de Souza. "Determinants of homicides in the state of Bahia, Brazil, in 2009." *Revista Brasileira de Epidemiologia* 17.1 (2014): 135-146.
- SOUSA, Geziel dos Santos de, et al. "Social determinants and their interference in homicide rates in a city in northeastern Brazil." *Revista Brasileira de Epidemiologia* 17 (2014): 194-203.