Genetics and Molecular Biology, 37, 1 (suppl), 186-193 (2014)

www.sbg.org.br

Research Article

From rumors to genetic isolates

Eduardo E. Castilla[1] and Lavinia Schuler-Faccini[2]

[1]*Instituto Nacional de Genética Médica Populacional, Latin American Collaborative Study of Congenital Malformations, Center for Medical Education and Clinical Research, Buenos Aires, Argentina.*
[2]*Departamento de Genética, Instituto Nacional de Genética Médica Populacional, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*

Abstract

Here we propose a registration process for population genetic isolates, usually geographic clusters of genetic disorders, based on the systematic search of rumors, defined as any type of account regardless of its reliability. Systematically ascertained rumors are recorded, and validated through a progressive process of pre-established steps. This paper outlines the conceptual basis for this approach and presents the preliminary results from a rumor-based nationwide registry of genetically isolated populations, named CENISO (Censo Nacional de Isolados), operating in Brazil since 2009. During the first four years of its existence (2009-2013), a total of 191 Rumors were registered and validated, resulting in a prevalence rate of one per million inhabitants of Brazil. When the five statutory geographic regions of Brazil were considered, more Rumors were registered for the Northeast (2.11; 1.74-2.54 per 10^6) than for the remaining four regions, North, Center-West, Southeast, and South, which did not differ among themselves. About half (86/191) of the recorded rumors were proven to be geographic clusters; of these disorders, 58 were autosomal recessive, 17 autosomal dominant, 5 X-linked, 3 multifactorial, and one environmental (thalidomide embryopathy).

Keywords: rumors, disease clustering, genetic diseases, population registers, Brazil.

Introduction

Rumors, clusters, alarms, and endemics or epidemics constitute a semi-continuous chain of events, some of which could be difficult to categorize, particularly those on the border between two categories. Working definitions for these different terms in the ECLAMC (Latin American Collaborative Study of Congenital Malformations) have been published elsewhere (Castilla and Orioli, 2004). More authorized significances can be found in Last's dictionary of epidemiology (Porta, 2008).

Since 1967, the ECLAMC program has conducted birth defect surveillance aimed at the detection and investigation of unusual occurrences in time and/or space. For time clusters, or epidemics, routine monitoring is performed, and quarterly data are compared against other equivalent surveillance systems through the International Clearinghouse for Birth Defects Surveillance and Research (ICBDSR, 2011). In addition, Focus, a continuous protocol for the study of space clusters, or endemics, based on the systematic evaluation of rumors, has been on-going since 1967 in all South-American countries, including Brazil (Castilla and Orioli, 2004).

The complete and systematic survey of genetically isolated populations and related geographic clusters of genetic disorders has been performed for small and well-defined populations, such as those in Finland (de la Chapelle, 1993), Israel (Zlotogora *et al.*, 2009), and the American Old Order Amish (Strauss and Puffenberger, 2009). Conversely, a similar study of the population of a large country has never been attempted to the authors' knowledge.

With a population of close to 200 million inhabitants, the population of Brazil makes up one half of South America, and one third of Latin America. The genetic structure of Brazilian populations had been thoroughly investigated since the late 1950's through the study of the frequency of consanguineous marriages in Roman Catholic Church records by Newton Freire-Maia (Freire-Maia, 1958; Freire-Maia and Freire-Maia, 1961), as well as by the quantification of the ethnic admixture from biological markers in northeastern Brazilian populations by Newton Morton and Henrique Krieger (Morton, 1964; Krieger *et al.*, 1965), among other methodologies. Furthermore, geographic clusters of some single gene diseases in Brazil have been known about for many decades. The pioneer publications include the studies of new mutations for Grebe's achondrogenesis (OMIM #200700) in the state of Bahia (Quel-

Send correspondencd to Lavinia Schuler-Faccini. Departamento de Genética, Instituto Nacional de Genética Médica Populacional, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brazil. E-mail: lavinia.faccini@ufrgs.br.

ce-Salgado, 1964); acheiropodia (OMIM #200500) in the state of Minas Gerais (Freire-Maia, 1975, 1981); and oculo-cutaneous albinism (OMIM #203200) in Ilha dos Lençóis (Freire-Maia *et al.*, 1978). Nevertheless, no systematic survey of genetically isolated populations, and/or geographic clusters of genetic disorders, has ever been performed in Brazil on a nationwide level.

In 2008, the Brazilian government created the 'Instituto Nacional de Genética Médica Populacional (INAGEMP)' for the study of medical genetics at the population level. Within the framework of this institute, a specific program for the study of genetically isolated populations was started in 2009 under the name CENISO, a Portuguese acronym for National Census of Isolates: 'Censo Nacional de Isolados'. CENISO is a surveillance system of sub-populations with endemic chronic diseases, most of them Mendelian disorders, aimed at their study, needs assessment, prevention and care.

This work presents a method for searching for genetic isolates and geographic clusters of genetic conditions based on the systematic collection, recording and validation of any rumors, regardless their source and reliability, as well as the results of its application in Brazilian populations.

Materials and Methods

Data collection

The data included here correspond to the first 48 months of operation of the CENISO program, from its start in August 2009 until the date of the last rumor included in this report: July 31, 2013.

Geographic distribution

Brazil is officially subdivided into more than 5,500 counties or municipalities, which are grouped into 27 States, and assembled into 5 Regions. Denominators for local population size were obtained from the Brazilian Institute of Geography and Statistics (IBGE) (Brazilian Institute of Geography and Statistics, 2013), and expected values for disease prevalence rates from the Unique Health System (SUS) (DATASUS, Brazilian Universal Health Service, 2013). Due to the limited sample size of less than two hundred observations, only Regions were considered for grouping localities in this work. Each locality was defined by their geographic coordinates at the degree and minute level. For space areas, such as municipalities or states, coordinates of their capital or administrative head localities were used.

Illnesses and their etiologies

Illnesses were defined by their OMIM code number www.omim.org whenever possible. Furthermore, ICBDX-BPA (International Classification of Diseases, Tenth edition, extended to a fifth code digit by the British Paediatric Association), was used when a specific coding slot was available.

Definitions

Rumors (gossip, hearsay) are any type of account, oral or written, of the unusual occurrence of a given fact, which in our case would be a suspected isolated population, and/or the unusual occurrence of a genetic or malformation disorder.

A geographic cluster is defined by a prevalence rate higher than expected (as determined from comparable population data) for a given disorder, in a population living in a defined geographic area, over a long period of time. However, for genetic diseases, this definition was reformulated, as the diseases in question are usually very rare, and expected prevalence rates are commonly unknown. Because of asymptomatic heterozygotes for recessive conditions, and non-penetrants for dominant conditions, what actually needs to be considered is the frequency of genes and genotypes rather than that of phenotypes (Castilla and Orioli, 2004)

Rumor Collection: The core of this program is the regular ongoing collection of rumors, as defined above, in a given population. The search for these rumors is proactive, by disseminating the question: "do you know any population with genetic problems?" This direct, though loose, question was initially disseminated through two main channels: national genetics meetings, and the internet. The former included the distribution of a simple reporting form with the collaboration of the Brazilian Society of Genetics, and the Brazilian Society of Medical Genetics, while the latter is carried out through the availability of open reporting access to the INAGEMP webpage http://www.INAGEMP.bio.br.

Rumor Validation: As rumors are groundless by definition, and most of them will prove to be false, validation efforts should be proportional to the reliability of the rumor. Registration is performed without any exclusion criteria, and a validation process, which is structured in four progressive phases, is then applied to registered rumors.

Phases of the study

Phase I is the registration of the Rumor, which is entered into the rumor registry. This is open to public observation at www.INAGEMP.bio.br through the link CENISO, and the button 'PARA VER AS POPULAÇÕES JÁ REGISTRADAS (pdf)': 'To See Already Registered Populations (pdf).

Phase II is the definition of the rumor. A simple one-page form is sent to the reporting person, for the recording of name, birth year, birth place, and denomination of the anomaly type for each known affected individual in the suspected Cluster. Denominators for the reported sample are then estimated from statutory databases according to a given space-time framework for the expected incidence

rate for the reported condition. If the rumor is substantiated by the provided information, Phase III is then initiated.

Phase III includes further definition and delineation of the rumor, through a brief site-visit. One or two medical geneticists from the CENISO staff visit the involved population with these three objectives: 1) to confirm the cluster, 2) to observe the population general conditions *in loco*, 3) to establish local contact with persons and institutions. A recording form with basic information is filled out for later discussion with the CENISO staff. This form includes information about the involved disease details, and diagnostic certainty, local resources, including hospitals, day care centers, physicians, nurses, and social workers, parochial and civil registration books, local natural leaders, community perception of the problem, proposed strategies and action plans. Due to the long distances usually involved in travelling in a country as large as Brazil, there are reserved funds at INAGEMP to finance these campaigns.

Phase IV is the development of a research project, if justified. If the cluster is already being studied by other research groups, collaboration and/or support is offered if needed. Community aspects are discussed with the leaders and support is provided after needs are assessed.

Statistical analysis

Confidence intervals at a critical value of 5% were used in this study.

Ethics considerations

Data included here refer to identified human subpopulations in which all individual human subjects were anonymized at the initial registration phase. Brazilian legislation (Resolução CNS 466/2012) does not require IRB approval for data obtained from public databases, as is the case for CENISO.

Results

After excluding three repeated and 13 irrelevant entries, 191 of 207 reported rumors were registered in the CENISO database during the initial 48 months of operation. A summary of the whole database is available at the INAGEMP-CENISO webpage, http://www.INAGEMP.bio.br/sis/produtos/CENISO_planilha.pdf#.

Figure 1 illustrates the geographic distribution of the 191 reported rumors, 86 of which were already confirmed as clusters, in the five official geographic regions of Brazil, which is further analyzed in Table 1.

While the observed total prevalence rate of reported rumors was one per million inhabitants (0.989; 95% CI 0.854-1.140), the northeastern Region (NE), had a significant higher prevalence rate of rumors (2.110; CI: 1.740-2.540) than the other four regions, north (N), center-west (CW), southeast (SE), and south (S), which did not differ among themselves. Furthermore, this difference was also true for clusters, for which NE had a higher prevalence

Figure 1 - Map of Brazil subdivided in 5 regions showing 191 reported rumors (Dots), 86 of which have already been proven as clusters (Squares). The five Official Regions of Brazil are: North, Northeast, Center-West, Southeast, and South.

(1.017 per million; CI: 0.764-1.330) than the remaining four regions (See Table 1).

The reported 191 rumors represented 86 proven geographic clusters, involving 58 autosomal recessive disorders, 17 autosomal dominant disorders, 5 X-linked disorders, 3 multifactorial disorders, and 1 environmental disorder. The environmental disorder is thalidomide embryopathy, resulting from a drug used for the treatment of leprosy in endemic areas of Brazil (Vianna *et al.*, 2013). Autosomal recessive disorders were the most prevalent rumors [0.534 per million inhabitants (CI: 0.435-0.647)] and clusters [0.300 per million inhabitants (CI: 0.228-0.388)] compared with the remaining etiologic groups: autosomal dominant, X-linked, multifactorial, environmental, and unspecified etiology (Table 2).

Likewise, the NE Region had a significant higher prevalence rate of rumors for autosomal recessive disorders per million inhabitants (1.280; CI: 0.995-1.620) than the other four regions. Nevertheless, this difference could not be substantiated for proven clusters (Table 3).

The 86 identified clusters involve 70 different disorders, with 16 instances of repetition of the same condition

Table 1 - Regional distribution of 191 reported rumors and 86 clusters.

Region	Inhabitants	Rumours			Clusters		
	Nº	Nº	Rate /1,000,000	95% CI	Nº	Rate /1,000,000	95% CI
N: north	16,318,163	7	0.429	0.172-0.884	2	0.123	0.015-0.443
NE: northeast	53,081,510	112	2.110	1.740-2.540	54	1.017	0.764-1.330
CW: center-west	14,423,952	3	0.208	0.043-0.608	3	0.208	0.043-0.608
SE: southeast	81,565,983	44	0.539	0.392-0.724	18	0.221	0.131-0.349
S: south	27,665,289	23	0.831	0.527-1.250	8	0.289	0.125-0.570
Geographically disperse		2			1		
Total	193,054,897	191	0.989	0.854-1.140	86	0.445	0.356-0.550

Table 2 - Etiologic grouping of 191 reported rumors and 86 clusters.

Etiology	Rumors			Clusters		
	Nº	Rate /1,000,000	95% CI	Nº	Rate /1,000,000	95% CI
AR	103	0.534	0.435-0.647	58	0.300	0.228-0.388
AD	40	0.207	0.148-0.282	17	0.088	0.051-0.141
XL	9	0.047	0.021-0.088	5	0.026	0.001-0.060
MF	16	0.083	0.047-0.135	3	0.016	0.003-0.045
ENV	9	0.047	0.021-0.088	1	0.005	0.001-0.028
NS	14	0.073	0.039-0.122	2	0.010	0.001-0.037
Total	191	0.989	0.854-1.140	86	0.445	0.356-0.550

Prevalence rates over 193,054,897inhabitants.
Autosomal recessive: AR disorders, autosomal dominant: AD, x-linked: XL, multifactorial: MF, environmental: ENV, not-specified: NS.

Table 3 - Regional distribution of 103 reported rumors and 58 clusters of autosomal recessive disorders.

Region	Inhabitants	Rumors			Clusters		
	Nº	Nº	Rate /1,000,000	95% CI	Nº	Rate /1,000,000	95% CI
N: north	16,318,163	1	0.061	0.002-0.341	1	0.061	0.002-0.341
NE: northeast	53,081,510	68	1.280	0.995-1.620	37	0.069	0.049-0.096
CW: center-west	14,423,952	3	0.208	0.043-0.608	3	0.208	0.043-0.608
SE: southeast	81,565,983	25	0.307	0.198-0.452	14	0.172	0.094-0.288
S: south	27,665,289	6	0.217	0.080-0.472	2	0.072	0.009-0.261
Total	193,054,897	103	0.534	0.435-0.647	58	0.300	0.228-0.388

clustered in more than one geographic location, none of them contiguous (Table 4). Fifty one of the 70 identified diseases are autosomal recessive traits, 12 are autosomal dominant, 5 are X-linked recessive, and 2 are unknown.

The 13 disorders involved in more than one cluster are very rare conditions. All but one produce autosomal recessive phenotypes, with expected gene prevalence rates below one percent in the population at large. These conditions are (MIM codes in parenthesis): Spinocerebellar Ataxia Type 3 (109150), Albinism, oculocutaneous (203100), Fraser Syndrome (219000), Gangliosidosis Type I (230500), Meckel Syndrome (249000), Cartilage hair hypoplasia (250250), Mucopolysaccharidosis Type IV: Morquio (253000), Spinal Muscular Atrophy Type I (253300), Muscular Dystrophy, Limb-Girdle type (253601), Deafness, Connexin 26 type (605428), Charcot-Marie-Tooth (606482), Mucopolysaccharidosis Type I: Hurler (607014), and Diaphanospondylodysostosis (608022).

Discussion

To the authors' best knowledge, a nationwide, systematic register of genetic clusters, such as CENISO in

Table 4 - Description of 86 identified clusters, ordered by phenotypic condition and space.

ID	Phenotypic condition			Space		
	Name	MIM	ICD10 /BPA	Locality name	IBGE code	Coordinates Lat S - Lon W
042	“Yellow people” undefined syndrome		R17	Craibas	2702355	9°37’-36°46’
062	Oral Clefts		Q37	Alfenas	3101607	21°25’-45°56’
036	Achondroplasia	100800	Q77.4	Sao Miguel	2412500	6°12’-39°29’
026	Apert Syndrome	101200	Q87.01	Paulinia	3536505	22°45’-47°09’
004	Aniridia	106210	Q13.1	Agua Branca	2700102	09°15’-37°56’
017	Spinocerebellar Ataxia, Type 3	109150	G11.8	Geographically dispersed, Rio Grande do Sul state	43	30° 02’-51° 13’
072	Spinocerebellar Ataxia, Type 3	109150	G11.8	Jerico	2507408	6°32’-37°48’
010	Fish-Eye Disease	136120	Q13.9	Betania do Piaui	2201739	08°08’-40°47’
023	Thalidomide embryopathy	142900	Q86.82	Cajarí	2102507	03°19’-45°00’
077	Huntington Disease	143100	G10	Feira Grande	2702603	9°54’-36°40’
015	Li-Fraumeni	151623	Q87.8	Geographically dispersed, Parana State	41	25°26’-49°16’
057	Myotonic Dystrophy, DM1	160900	G71.11	Santa Cruz	2513208	6°31’-38°03’
078	Spinocerebellar Ataxia, Type 1	164400	G11.8	Sao Paulo	3550308	23°29’-46°38’
079	Spinocerebellar Ataxia, Type 3	164500	G11.8	Rio de Janeiro	3304557	22°49’-43°12’
006	Rheumatoid Arthritis	180300	I00.	Mangueirinha	4114401	25°26’-52°10’
005	Acheiropody	200500	Q73.8	Minas Gerais, State	31	19° 55’-43° 57’
001	Chondrodysplasia, Grebe type	200700	Q77.00	Bahia, State	29	12° 56’-38° 31’
076	Congenital Adrenal Hyperplasia	201910	E25.0	Gado Bravo	2506251	7°34’-35°47’
008	Adrenocortical Carcinoma, Hereditary, ADCC	202300	C74.0	Curitiba	4106902	25°26’-49°16’
037	Albinism, oculocutaneous	203100	E70.3	Santana do Mundau	2708105	9°10’-36°13’
038	Albinism, oculocutaneous	203100	E70.3	Salvador (Ilha Mare)	2927408	12°58’-38°02’
039	Albinism, oculocutaneous	203100	E70.3	Miguel Calmon	2921203	11°25’-40°35’
040	Albinism, oculocutaneous	203100	E70.3	Marau	2920700	14°06’-38°59’
003	Albinism	203200	E70.3	Curupuru	2103703	01°49’-44°51’
044	Ataxia Telangectasia	208900	G11.30	Sao Francisco	2513984	6°36’-38°05’
071	Seckel Syndrome	210600	Q87.18	Sto Antonio Posse	3548005	22°36’-46°55’
086	Cerebrotendinous xanthomatosis	213700	E75.50	Queimadas	2512507	7°21’-35°54’
007	Chondrodysplasia, Bloomstrand type	215045	Q78.	Mata Grande	2705002	09°07’-37°44’
068	Fraser Syndrome	219000	Q11.2	Itu	3523909	23°15’-47°17’
069	Fraser Syndrome	219000	Q11.2	Tres Lagoas	5008305	20°45’-51°41’
052	Dandy-Walker	220200	Q03.1	Indaiatuba	3520509	23°05’-47°13’
045	Friedreich Ataxia	229300	G11.11	Piloes	2410009	6°16’-38°02’
073	Gangliosidosis, Type I	230500	E75.10	Porto Alegre	4314902	30°1’-51°13’
075	Gangliosidosis, Type I	230500	E75.10	Jundiai	3525904	23°11’-46°53’
074	Gaucher Disease, Type I	230800	E75.21	Tabuleiro do Norte	2313104	5°15’-38°07’
012	Glycogen Storage Disease, Ia	232200	E74.01	Caxias do Sul	4305108	29°06’-51°11’
061	Maple Syrup Urine Disease, MSUD	248600	E71.0	Vinhedo	3556701	23°01’-46°58’
063	Meckel Syndrome	249000	Q61.9	Itape	2916203	14°53’-39°25’
064	Meckel Syndrome	249000	Q61.9	Piracicaba	3538709	22°43’-47°38’
048	Cartilage hair hypoplasia	250250	Q78.8	Jequitinhonha	3135803	16°26’-41°00’
049	Cartilage hair hypoplasia	250250	Q78.8	Campinas	3509502	22°54’-47°03’
065	Mucolipidosis, Type II	252500	E77.00	Girau do Ponciano	2702900	9°53’-36°49’
032	Mucopolysaccharidosis, Type IIIC: Acetyl CoA	252930	E76.23	Serra Branca	2504702	7°28’-36°39’
018	Mucopolysaccharidosis, Type IV A: Morquio	253000	E76.25	Serra Branca	2515500	07°28’-36°39’

Table 4 - cont.

ID	Phenotypic condition			Space		
	Name	MIM	ICD10 /BPA	Locality name	IBGE code	Coordinates Lat S - Lon W
033	Mucopolysaccharidosis, Type IV A: Morquio	253000	E76.25	Mombaça	2308500	5°44'-39°37'
034	Mucopolysaccharidosis, Type IV A: Morquio	253000	E76.25	Campina Grande	2504009	7°12'-35°52'
019	Mucopolysaccharidosis, Type VI: Maroteaux	253200	E76.28	Monte Santo	2921500	10°26'-39°19'
035	Mucopolysaccharidosis, Type VI: Maroteaux	252650	E76.28	Quixere	2311504	5°4'-37°59'
043	Spinal Muscular Atrophy, Type I	253300	G12.1	Vieiropolis	2517209	6°30'-38°15'
046	Spinal Muscular Atrophy, Type I	253300	G12.1	Sao Miguel	2412500	6°12'-39°29'
055	Muscular Dystrophy, Limb-Girdle	253601	G71.08	Ouro Branco	2408508	6°42'-36°56'
056	Muscular Dystrophy, Limb-Girdle	253601	G71.08	Jerico	2507408	6°32'-37°48'
020	Neu-Laxova	256520	Q04.9	PousoAlegre	3152501	22°13'-45°56'
067	Opsismodysplasia	258480	Q78.8	Belo Jardim	2601706	8°20'-36°25'
021	Osteogenesis imperfect, Type 3	259420	Q78.00	BuenoBrandao	3109105	22°26'-46°21'
009	Isolated Growth Hormone Deficiency	262400	E23.01	Itabaianinha	2803005	11°16'-37°47'
085	Short Limb Polydactyly III: Verma-Namouff	263510	Q77.2	Gameleira	2605905	8°35'-35°23'
070	Postaxial Acrofacial Dysostosis	263750	Q75.1	Sumare	3552403	22°49'-47°15'
016	Lipodystrophy, Congenital Generalized	269700	E88.1	Sao Miguel	2412500	06°12'-39°29'
011	Twinning	276400	P01.5	CândidoGodói	4304309	27°57'-54°45'
081	Usher Syndrome	276900	H35.50	Sao Miguel	2412500	6°12'-39°29'
025	Xeroderma Pigmentosum	278730	Q82.1	Faina	5207535	15°26'-50°21'
080	Lesh-Nyhan Syndrome	300322	E79.1	Sao Miguel	2412500	6°12'-39°29'
059	Muscular Dystrophy, Becker Type	300376	G71.00	Ouro Velho	2510600	7°17'-37°09'
041	Alpha-thalassemia, mental retardation	301040	D56.0	Sao Miguel	2412500	6°12'-39°29'
031	Mucopolysaccharidosis, Type II: Hunter	309900	E76.1	Aquiraz	2301000	3°54'-38°23'
060	Muscular Dystrophy, Duchenne Type	310200	G71.06	Sao Miguel	2412500	6°12'-39°29'
027	Laron Syndrome	600946	E34.32	Orobo	2609709	7°44'-35°36'
024	Tricoepithelioma, multiple familiar	601606	Q84.9	Aracati	2301109	04°33'-37°46'
013	Cerebello-trigeminal-dermal dysplasia	601853	Q87.8	Ribeirao Preto	3543402	21°10'-47°48'
047	Breast cancer, familial	604370	C50.9	Geographically dispersed-South Region		
082	Deafness (Connexin 26)	605428	H91.9	Queimadas	2512507	7°21'-35°54'
084	Deafness	605428	H91.9	Maracana	1504307	0°35'-47°31'
083	Deafness (DFNA18)	606012	H91.9	Sao Miguel	2412500	6°12'-39°29'
050	Charcot-Marie-Tooth	606482	G60.00	Sao Miguel	2412500	6°12'-39°29'
051	Charcot-Marie-Tooth	606482	G60.00	Sossego	2516151	6°45'-36°14'
029	Mucopolysaccharidosis, Type I: Hurler	607014	E76.0	Piloes	2410009	6°16'-38°02'
030	Mucopolysaccharidosis, Type I: Hurler	607014	E76.0	Jangada	5104906	15°14'-56°29'
066	Niemann-Pick Disease, Type B	607616	E75.25	Santa Cruz	2513208	6°31'-38°03'
058	Muscular Dystrophy, Congenital	607855	G71.25	Piloes	2410009	6°16'-38°02'
053	Diaphanospondylodysostosis	608022	Q78.8	Itupiranga	1503705	5°8'-49°19'
054	Diaphanospondylodysostosis	608022	Q78.8	Campinas	3509502	22°54'-47°03'
022	SPOAN syndrome	609541	Q87.8	Serrinha dos Pintos	2413557	06°06'-37°57'
014	Ichthyosis, Congenital	612281	Q80.9	Humaita	4309704	27°33'-53°58'
028	Spastic Paraplegy	612319	G11.4	Sao Miguel	2412500	6°12'-39°29'
002	Santos Syndrome	613005	Q87.8	Riacho de Santana	2410801	06°15'-38°19'

ID: Cluster Identification code. MIM: Mendelian Inheritance in Man code. ICD10/BPA: ICD10 with British Paediatric Association fifth digit.
IBGE code: Brazilian Institute of Geography and Statistics municipality code, in which the first two digits identify the State (http://www.ibge.gov.br/home/geociencias/areaterritorial/area.shtm).

Brazil, is an almost unique program, even considering some similar, albeit not equal, ongoing programs in other parts of the world, such as the aforementioned studies of Amish and Mennonites in the US and Canada (Strauss and Puffenberge, 2009; Rider *et al.*, 2011), Arabs in Israel (Zlotogora *et al.*, 2009), and isolates in Finland (de la Chapelle, 1993).

The purpose of CENISO is to base the building of a registry on sensitive and imprecise rumors, provided that validation by a cost-benefit efficient process is used. This system could be applied to other populations, not only for the registration and surveillance of endemic areas but also for outbreaks of diseases caused by environmental agents.

About half (86/191: 45.0%) of the CENISO-registered rumors were confirmed as geographic clusters, while one third (57/191: 29.8%) has not yet reached validation phase-2 for Rumor definition, and therefore are not yet confirmed. While low specificity is expected for this searching approach, based on extensive recording of unselected anecdotes, this low validation rate could also reflect a passive response rate from the CENISO coordination during the reported period. Nevertheless, unlike time/space clusters of diseases, also known as epidemics (Williams *et al.*, 2002), most genetic etiologic factors are rather stable over time, undergoing gradual variations through generations, mainly due to cultural changes in reproductive patterns within genetic population isolates. Existence of space clusters over such a long period reasonably explains their higher reliability when compared with time/space clusters (Williams *et al.*, 2002).

More reported rumors were located in the Northeastern region than in the rest of Brazil, a finding which is most likely related to the reported higher prevalence of consanguineous marriages in this region (Freire-Maia, 1958; Krieger *et al.*, 1965; Weller *et al.*, 2012). With only approximately 15 million inhabitants each, two of the five Brazilian geographic regions, North and Central-West are markedly less populated then the Northeast, with a population of 53 million. However, the low concentration of clusters in the remaining two regions cannot be explained by small population size, namely, the South with 28, and the Southeast with 82 million residents.

Because most genetically isolated sub-populations, for geographic or cultural reasons, are also inbred, the observed higher involvement of autosomal recessive phenotypes is expected. Nevertheless, autosomal dominant and X-linked, as well as oligo or polygenic traits, are also observed in the registered clusters in spite of not being directly related to inbreeding. One possibly associated factor in these situations is the readiness to recognize familial aggregation in small, defined populations where acquaintance of relatives and knowledge of ancestors through several generations is more complete compared to families living in urban large cities. Another explanation is the low mobility of isolated populations, so large pedigrees with dominant or X-linked mutations are concentrated on small geographic areas. The aniridia cluster in Alagoas is an example of this phenomenon (Fernandes-Lima *et al.*, 2013). Furthermore, small isolates could have randomly selected specific mutations during successive genetic drifts starting from a small number of founders and going through reductions on the effective population size in periods of high mortality in the past, as reported for the Finnish population (de la Chapelle, 1993).

More recently, Moreau *et al.* (2011) have shown a selective advantage for people in the wave front of population range expansions in Quebec, Canada. This differential reproduction in wave front migrating individuals can lead to substantial changes in gene frequency in the derived populations. Considering the historical patterns of colonization and internal migrations in Brazil, we could expect that phenomena similar to those seen in Canada might have happened here as well. The cancer clusters in southern Brazil (Achatz *et al.*, 2009) could be an example of such a phenomenon, but further investigation is needed to test this hypothesis.

One drawback of this report is our ignorance of the completeness of ascertainment of clusters in Brazil. This will most likely come with time as the current register continues operating, with the essential collaboration from the users, as stated for the Israeli National Genetic Database by Zlotogora *et al.* (2009).

Expansion of this registry of rumors and clusters to the rest of South-America, first, and the whole of Latin America, later, is planned at CENISO after the Brazilian registry is consolidated. This extension to other countries will mainly, though not exclusively, use the ECLAMC hospital network, an INAGEMP program described elsewhere (Castilla and Orioli, 2004).

Acknowledgments

We thank Juan Gilli and Marcelo Zagonel de Oliveira for their participation in the geo-reference analysis of the reported data. To Gabriela and Vanessa Paixão-Cortes for their participation in organizing clusters and rumors spreadsheets. Financial support was provided by the following agencies: INCT-INAGEMP; Ministry of Science and Technology/CNPq, (grants no. 476978/2008-4 and 306750/2009-0), a FIOCRUZ-FIOTEC fellowship to EEC, a CONICET, National Research Council of Argentina contract as researcher to EEC, a CNPq, research fellowship to LSF.

References

Achatz MI, Hainaut P and Ashton-Prolla P (2009) Highly prevalent TP53 mutation predisposing to many cancers in the Brazilian populations: Case for newborn screening? Lancet Oncol 10:920-925.

Castilla EE and Orioli IM (2004) ECLAMC: The Latin American collaborative study of congenital malformations. Community Genet 7:76-94.

de la Chapelle A (1993) Disease gene mapping in isolated human populations: The example of Finland. J Med Genet 30:857-865.

Fernandes-Lima ZS, Paixão-Côrtes VP, Andrade AKM, Fernandes AS, Coronado BNL, Monte-Filho HP, Santos MJ, Omena-Filho RL, Biondi FC, Ruiz-Linares A, *et al.* (2013) Ocular and craniofacial phenotypes in a large Brazilian family with congenital aniridia. Clin Genet (Epub ahead of print).

Freire-Maia N (1958) Consanguineous marriages in Brazil. I -Structure of such marriages. II. Factors of geographic distribution. Eugen Quart 5:105-114.

Freire-Maia A (1975) Genetics of acheiropodia (the handless and footless families of Brazil). Clin Genet 7:98-102.

Freire-Maia A (1981) Historical note: the extraordinary handless and footless families of Brazil - 50 years of acheiropodia. Am J Med Genet 9:31-41.

Freire-Maia N and Freire-Maia A (1961) The structure of consanguineous marriages and its genetic implications. Ann Hum Genet 25:29-39.

Freire-Maia N, de Andrade FL, de Athayde-Neto A, Cavalli IJ, Oliveira JC, Marçallo FA and Coelho A (1978) Genetic investigations in a Northern Brazilian island. II. Random drift. Hum Hered 28:401-410.

ICBDSR (2011) Annual Report, 2011 with Data for 2009. Centre of the International Clearinghouse for Birth Defects Surveillance and Research of ICBDSR, Rome, 302 pp.

Krieger H, Morton NE, Mi MP, Azevedo E, Freire-Maia A and Yasuda N (1965) Racial admixture in northeastern Brazil. Ann Hum Genet 29:113-125.

Moreau C, Bhérer C, Vézina C, Jomphe M, Labuda D and Excoffier L (2011) Deep human genealogies reveal a selective advantage to be on an expanding wave front. Science 334: 1148-1150.

Morton NE (1964) Genetic studies of Northeastern Brazil. Cold Spring Harbor Symp Quant Biol 29:69-79.

Porta M (2008) A Dictionary of Epidemiology. 5th edition, Oxford University Press, New York, 320 pp.

Quelce-Salgado A (1964) A new type of dwarfism with various bone aplasias and hypoplasias of the extremities. Acta Genet Stat Med 14:63-66.

Rider NL, Strauss KA, Brown K, Finkenstedt A, Puffenberger EG, Hendrickson CL, Robinson DL, Muenke N, Tselepis C, Saunders L, *et al.* (2011) Erythrocyte pyruvate kinase deficiency in an old-order Amish cohort: longitudinal risk and disease management. Am J Hematol 86:827-834.

Strauss KA and Puffenberger EG (2009) Genetics, medicine, and the Plain people. Annu Rev Genomics Hum Genet 10:513-536.

Vianna FSL, Schüler-Faccini L, Leite JCL, de Sousa SHC, da Costa LMM, Dias MF, Morelo EF, Doriqui MJR, Maximino CM and Sanseverino MTV (2013) Recognition of the phenotype of thalidomide embryopathy in countries endemic for leprosy: new cases and review of the main dysmorphological findings. Clin Dysmorphol 22:59-63.

Weller M, Tanieri M, Pereira JC, Almeida Edos S, Kok F and Santos S (2012) Consanguineous unions and the burden of disability: a population-based study in communities of Northeastern Brazil. Am J Hum Biol 24:835-840.

Williams LJ, Honein MA and Rasmussen SA (2002) Methods for a public health response to birth defects clusters. Teratology 66:S50-S58.

Zlotogora J, van Baal S and Patrinos GP (2009) The Israeli National Genetic Database. Isr Med Assoc J 11:373-375.

Internet Resources

DATASUS, Brazilian Universal Health Service (2013) http://www2.datasus.gov.br/DATASUS/index.php?area = 0205&VObj=http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sinasc/cnv/nv (November 12, 2013).

IBGE, Brazilian Institute of Geography and Statistics, http://www.ibge.gov.br/home/estatistica/populacao/registrocivil/2011/default _xls.shtm (November 12, 2013).