Tuberculosis and diabetes: probabilistic linkage of databases to study the association between both diseases

Ricardo Gadelha de Abreu[1]
Artur Iuri Alves de Sousa[2]
Maria Regina Fernandes de Oliveira[3]
Mauro Niskier Sanchez[4]

[1]Ministério da Saúde, Secretaria de Vigilância em Saúde, Brasília-DF, Brasil
[2]Ministério da Saúde, Agência Nacional de Vigilância Sanitária, Brasília-DF, Brasil
[3]Universidade de Brasília, Núcleo de Medicina Tropical, Brasília-DF, Brasil
[4]Universidade de Brasília, Departamento de Saúde Coletiva, Brasília-DF, Brasil

Abstract

Objective: to describe the profile of cases of tuberculosis and diabetes comorbidity in Brazil. **Methods**: this is a descriptive study with data from the Brazilian Information System for Notifiable Diseases – tuberculosis (Sinan-TB) and from the System of Registration and Monitoring of Hypertension and Diabetes Mellitus (Hiperdia), from 2007 to 2011; probabilistic linkage was carried out with Reclink software. **Results**: 24,443 cases of comorbidity were found, including 3,181 cases not registered on Sinan-TB; of the total number of recovered cases, mostly were males (57.2%), aged 40-59 years (52.3%), black/brown-skinned (68.4%), with five to eight years of schooling (78.4%), with no regular use of alcohol (86.5%) and negative serology for the HIV virus (91.8%). **Conclusion**: the cases found had similar profile to those registered on Sinan-TB and the probabilistic linkage of data from different information systems enabled the detection of cases not captured by surveillance.

Keywords: Tuberculosis; Diabetes Mellitus; Information Systems; Epidemiology, Descriptive.

Correspondence:
Ricardo Gadelha de Abreu - CA 09, Lotes 1, 3 e 5, Torre 3, Apto. 206, Lago Norte, Brasília-DF, Brasil. CEP: 71503-509
E-mail: ricardo.gadelha@gmail.com

Introduction

Tuberculosis and diabetes are Public Health issues of great relevance. The risk of a person with diabetes to develop tuberculosis can be 2.44 to 8.33 times bigger than the same risk for a person without diabetes.[1] However, no reason was found for the higher incidence of Koch's bacillus infection, which causes tuberculosis, among diabetics. The tuberculosis-diabetes comorbidity requires more complex attention and care, since diabetes may interfere with the metabolism of antituberculosis drugs.[2]

Tuberculosis is a disease of compulsory notification. The key measures to control the disease and interrupt its transmission chain are the diagnosis and the correct and timely treatment of cases of pulmonary tuberculosis.[3,4]

In 2014, 9.6 million people were estimated to have tuberculosis around the world: 5.4 million men, 3.2 million women and 1 million children. However, nearly 6 million new cases of the disease – less than two-thirds (62.5%) – were reported to the World Health Organization (WHO), suggesting that 37.5% of the cases had not been diagnosed or reported. For the same year, 1.5 million deaths from tuberculosis were estimated.[5]

> *Tuberculosis and diabetes are Public Health issues of great relevance. The risk of a person with diabetes to develop tuberculosis can be 2.44 to 8.33 times bigger than the same risk for a person without diabetes.*

According to WHO, in 2015, Brazil was one of the 22 countries which concentrated 80% of the global burden of tuberculosis, occupying the 18th position in the absolute number of new cases and the 22nd position regarding the incidence rate.[5] According to the Brazilian Ministry of Health, the incidence rate of tuberculosis declined 38.7%; and the mortality rate presented a reduction of 33.6%, from 2003 to 2012.[6,7]

The growth in the number of people with diabetes in the world, from 171 million cases in 2000 to an estimate of 440 million in 2030, may represent an additional factor hindering tuberculosis control.[2]

In 2012, the prevalence of diabetes was estimated at 26.4 million people, and was forecast at 40 million in 2030 for the countries of Central and South America. The prevalence of diabetes in Brazil was estimated at 4.6% in 2000 (8th place among these countries). The estimate for 2030 points to a prevalence of 11.3% (6th place) in 2030.[8,9]

In Brazil, cases of tuberculosis are registered in the Brazilian Information System for Notifiable Diseases (Sinan). The registration and monitoring of people with diabetes in the Brazilian National Health System (SUS) were done through the System of Registration and Monitoring of Hypertension and Diabetes Mellitus (Hiperdia), up to 2013. Since then, Hiperdia was replaced by e-SUS Primary Health Care (PHC), or e-SUS PHC: a process of IT qualification, which is a strategy from the Department of Primary Health Care, Secretariat of Primary Health Care of the Ministry of Health, to restructure the Primary Health Care information offered by the SUS.[10-14]

The growing availability of systems capable of storing nationwide data requires the use of methodologies that connect these data to obtain bases for more robust epidemiological analyses.[15,16] The absence of a unique identifier in the databases to associate records from different health systems is an additional challenge.

To ensure the quality of the recording sources that assist the surveillance of tuberculosis, the linkage between databases can be justified by the qualification of not onerous and low operational costs information, enabling the detection of cases identified in other systems, but not captured by the disease surveillance.[17-19]

This study aimed to describe the profile of cases of tuberculosis and diabetes comorbidity in Brazil, from 2007 to 2011.

Methods

A descriptive study was conducted using data from cases of tuberculosis-diabetes comorbidity in Brazil, obtained from the probabilistic linkage of Sinan databases related to tuberculosis notifications (Sinan-TB), and Hiperdia, related to the records of diabetes cases. The database of Sinan-TB considered the cases reported from January 1st, 2001 to December 31st, 2012, and Hiperdia database considered cases from January 1st, 2007 to December 31st, 2011.

Sinan-TB is fed by the Tuberculosis Notification/Investigation Form, which includes information on the case, place of residence, clinical situation and classification according to the type of admission, among others. For epidemiological surveillance purposes, the National Program for Tuberculosis Control defined as notified case on Sinan-TB only the confirmed cases of the disease.[20]

Hiperdia was used for recording and monitoring individuals with diabetes and/or hypertension treated at SUS ambulatory care network. The system was created in 2002, and allowed to generate information that could subsidize the purchase, dispensing and distribution of drugs to people who were registered, and also contributed to effective planning of prevention and control of complications in individuals already diagnosed with the disease.[10-13] The record with single entry of the person diagnosed with diabetes at Hiperdia enabled their identification and the linkage to health facilities or teams of SUS Primary Health Care, who recorded the data, allowing the generation of information on the performance and clinical results during the follow-up.[21]

According to the Joint Ordinance No. 2 of the Secretariat of Health Surveillance and the Executive Secretariat, both from the Ministry of Health, dated March 5th, 2002, in its article 5, in order to integrate the National Pharmaceutical Assistance Program for Hypertension and Diabetes Mellitus, the municipality manager should sign a Commitment Form and be responsible for integrating the pharmaceutical assistance program, registering and monitoring ill individuals at Hiperdia, to ensure to the registered individuals the receipt of the prescribed drugs.[22]

The municipalities that joined the Pharmaceutical Assistance Program for Hypertension and Diabetes Mellitus were responsible for feeding Hiperdia national database, via registration of users in the Municipal Coordinating Subsystem or by exporting data from a local application.[13,23]

In order to minimize possible errors, prior preparation of the databases was conducted to exclude duplicate records and to standardize (use of the same spelling for the first syllables of names with the same phonetic) and encoding of fields. To improve the quality of the field 'name' used in the process, it was necessary to remove information or characters such as &, #, @, and others mistakenly added to the person's name.

In order to obtain a joint basis which covered the cases of tuberculosis and diabetes, and decreasing the number of comparisons between records, we used, in both databases, probabilistic procedures established by Reclink software, version 3.1.6.3160, using common fields to check if the matched records belonged to the same individual. The following routines to standardize common fields were adopted: (i) comparison between 'person's name', 'mother's name' and 'date of birth', and application of algorithms to approximate comparison of strings, with the objective to exclude phonetic errors and typos; (ii) blocking, by dividing the databases into mutually exclusive blocks, with comparisons restricted to the records of the same block and optimization of the comparison between records; (iii) calculation of scores to indicate the degree of agreement between records of the same pair; and (iv) setting limits for classification of pairs of records and identification as true or false pairs.[24,25]

Probabilistic procedures were applied to identify possible duplications at Hiperdia database. The duplicates removal routine was not applied on Sinan-TB database, because it had already been performed by the National Program for Tuberculosis Control, prior to the provision of data.

In order to speed the processing of the databases, we applied blocking in multiple steps, since they totaled more than one million records and the computers used did not have enough memory space and speed for processing this amount of data. Therefore, we divided Hiperdia database into six blocks – considering the first letter of the person's name – to compare to Sinan-TB database:

1st Names beginning from A to C
2nd Names beginning from D to I
3rd Names beginning from J to L
4th Names beginning with M
5th Names beginning from N to R
6th Names beginning from S to Z

The comparison of the records within each block followed three steps:

1) Comparison between the patient's full name, the patient's mother's full name and date of birth using the blocking key PBLOCO+UBLOCO+SEXO
2) Comparison between the patient's full name, the patient's mother's full name and date of birth using the blocking key PBLOCO+UBLOCO
3) Comparison between the patient's full name, the patient's mother's full name and date of birth using the blocking key PBLOCO+SEXO

Probabilistic linkage require greater attention to the dubious pairs located on the border of uncertainty, known as the gray zone.[26] For maximizing the tool's ability to identify the true pairs, linkage parameters were estimated with the available tool application Reclink – procedures 'generate matrix' and 'calculate parameters' – on the information of the databases themselves, generating a theoretical distribution of score: -1.6843 to 30.0763.

As the databases were too long, we used the sampling fraction of 1% in order to choose the best parameters in the linkage, i.e., the ability to better distinguish pairs from non-pairs. We checked the database, in order to identify the best limit for pairs rating: records with scores higher than or equal to 24.9 were considered true pairs and with scores lower than 24.9 were considered non-pairs or dubious pairs. With the estimated parameters, the performance in the identification of true pairs was quite good, which may suggest that the manual review of the gray zone would not add a substantial amount of records to the combined pairs. Thus, we decided not to review the gray zone and name dubious pairs of non-pairs.[27]

After processing the six blocks related to Sinan-TB database, the respective blocks were rejoined, resulting in a single database, whose names for the data sources were encoded as follows:

01 = Sinan-TB database

10 = Hiperdia database

11 = tuberculosis-diabetes comorbidity database (linked)

For the final database the codes 01 and 11 (respectively: tuberculosis cases without tuberculosis-diabetes comorbidity on Sinan from January 1st, 2007 to June 30th, 2011, and cases with tuberculosis-diabetes comorbidity – cases reported on Sinan-TB with the combination of information with diabetes and recovered cases after the linkage between Sinan-TB and Hiperdia). Cases of tuberculosis reported on Sinan-TB without associated diabetes were considered as no comorbidity, i.e. tuberculosis cases whose variable 'disease and associated diseases – diabetes' of Sinan Form presented code '2' (No); cases with code '9' (ignored) or unfilled (blank) were excluded for the same variable. For cases with comorbidity TB-diabetes, the following were considered:
- cases of tuberculosis notified on Sinan in which the variable 'disease and associated diseases - diabetes' presented the code '1' (Yes);
- cases of diabetes registered on Hiperdia before the case of tuberculosis was notified on Sinan; and
- cases of diabetes and tuberculosis registered and reported in the same year on both systems – Hiperdia and Sinan.

Cases of tuberculosis recorded on Sinan after June 2011 were excluded, since individuals diagnosed in the second half of 2011 should complete the recommended treatment (six months) only in 2012, period out of the study and with no information on diabetes records.

To describe the sociodemographic profile of individuals with tuberculosis-diabetes comorbidity, the following variables of Sinan-TB notification form were considered:
- sex (male, female);
- age groups (in years: up to 19; 20 to 39; 40 to 59; 60 or more);
- ethnicity/skin color (white, black/brown, other);
- education level (in years of schooling: up to 4, 5 to 8; 8 or more);
- regular alcohol intake (yes, no); and
- serology for human immunodeficiency virus, HIV (positive, negative).

The analyses were performed using Pearson's chi-square test to compare the proportions between the linkage group and Sinan-TB group, considering a level of significance of 5%.

Nominal databases of Sinan and Hiperdia were provided by the Ministry of Health, through the National Program for Tuberculosis Control and SUS IT Department (DATASUS), after the signature of a Statement of Responsibility by the researcher. The research project was approved by the Ethics Committee of the Faculty of Health Sciences, from the University of Brasilia: Report No 552,561, dated March 11th, 2014.

Results

After the databases organization, 1,090,375 records of tuberculosis cases and 1,246,137 diabetes cases were accounted for the period studied. (Figure 1)

In the first block, from A to C of Hiperdia database, 13,347 duplications were removed from the total of 209,370 entries. In the second block, from D to I of Hiperdia database, 15,497 duplications were removed from the total of 225,089 entries. In the third block, from J to L of Hiperdia database, 14,795 duplications were removed from the total of 238,660 entries. In the fourth block, first letter M of Hiperdia database, 20,387 duplications were removed from the total of 284,337 entries. In the fifth block, from N to R of Hiperdia database, 10,746 duplications were removed from the total of 146,261 entries. Finally, in the sixth block, from S to Z of Hiperdia database, 9,148 duplications were removed from the total of 142,412 entries. Altogether, 83,920 duplications were removed. Furthermore, from the total diabetes records (n=1,246,137), eight presented no valid information in the 'individual's name' field for the creation of the block, and were excluded from the analysis.
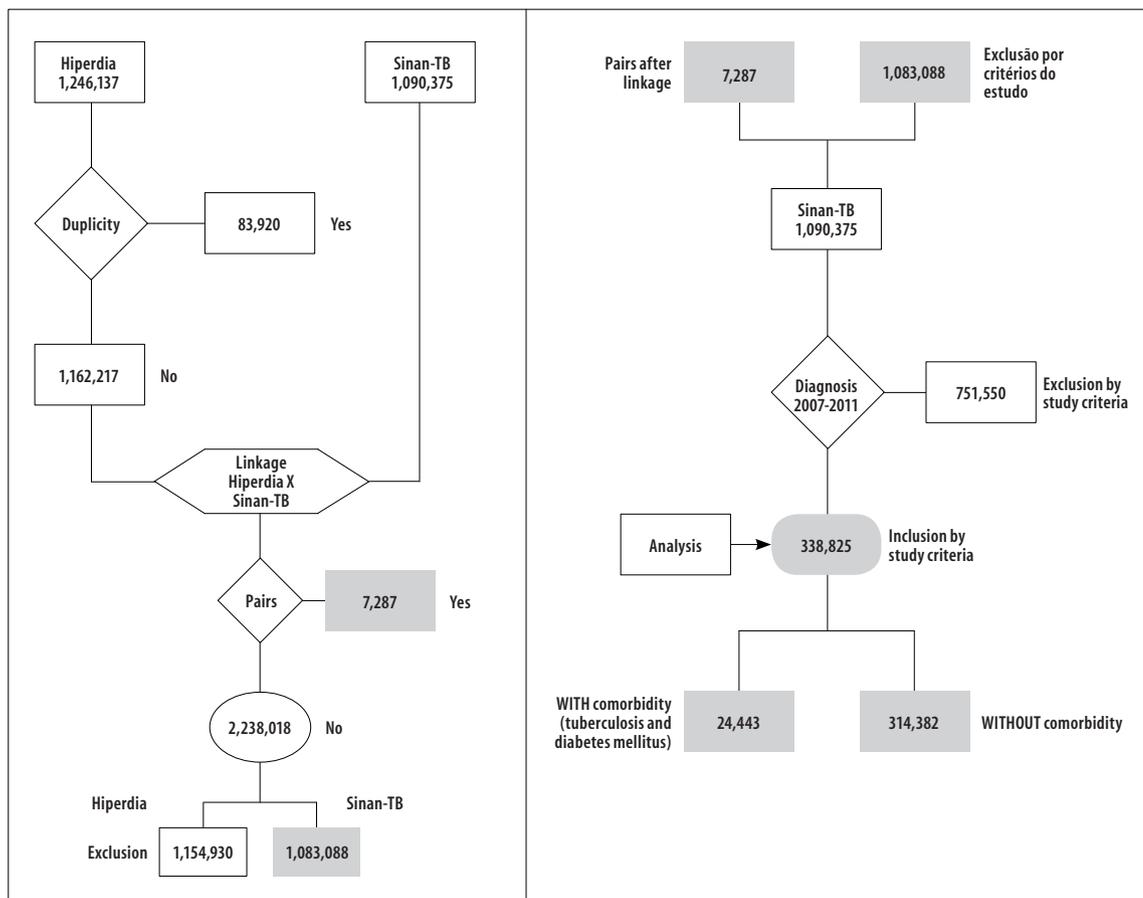
Subtitle:
Sinan-TB: Information System for Notifiable Diseases – tuberculosis
Hiperdia: System of Registration and Monitoring of Hypertension and Diabetes *Mellitus*

Figure 1 – Database flowchart for individuals with or without tuberculosis-diabetes *mellitus* comorbidity

After the linkage between Sinan-TB and Hiperdia databases, 7,287 records were found in both databases, i.e., with TB-diabetes comorbidity (Table 1):
- block 1 [A to C], 1,057 pairs;
- block 2 [D to I], 987 pairs;
- block 3 [J to L], 2,097 pairs;
- block 4 [M], 1,765 pairs;
- block 5 [N to R], 734 pairs;
- block 6 [S to Z], 647 pairs;

According to the data origin (Sinan-TB; Sinan-Hiperdia linkage) and considering the criteria for inclusion of cases for this study, the analysis gathered 335,644 cases of tuberculosis with or without comorbidity on Sinan and 3,181 cases of diabetes-tuberculosis comorbidity recovered in the linkage, i.e., which were not complete in the field of information on tuberculosis-diabetes association on Sinan-TB, totalizing 338,825 cases.

Of this total, 314,382 did not present comorbidity and 24,443 had diabetes and tuberculosis (21,262 from the Sinan-TB and 3,181 of the linked database) (Figures 1 and 2).

Among the cases of tuberculosis-diabetes comorbidity identified after the linkage, we observed a predominance of males (57.2%), 40-59 years old (52.3%), ethnicity/skin color black/brown (68.4%), five to eight years of schooling (78.4%), not regular alcohol intake (86.5%) and negative serology for HIV (91.8%). Among the cases of comorbidity reported on Sinan-TB database, the predominant characteristics were: males (63.1%), 40-59 years old (53.1%), ethnicity/skin color black/brown (56.1%), five to eight years of schooling (87.0%), not regular alcohol intake (84.9%) and negative serology for HIV (91.4%). There was no statistically significant difference (p >0.05) between the cases of comorbidity

Table 1 – Number of pairs found after linkage between the blocks of Sinan-TB[a] and Hiperdia standardized databases,[b] Brazil, 2007-2011

Blocks (initials of patients' names)	Total number of cases in the databases	Number of pairs found
A-C	Hiperdia = 195,172 Sinan-TB = 206,645	1,057
D-I	Hiperdia = 208,749 Sinan-TB = 206,840	987
J-L	Hiperdia = 222,301 Sinan-TB = 248,411	2,097
M	Hiperdia = 262,531 Sinan-tuberculose = 140,690	1,765
N-R	Hiperdia = 134,891 Sinan-tuberculose = 145,100	734
S-Z	Hiperdia = 132,736 Sinan-tuberculose = 135,498	647
Total of pairs		**7,287**

a) Sinan-TB: Information System for Notifiable Diseases – tuberculosis
b) Hiperdia: System of Registration and Monitoring of Hypertension and Diabetes Mellitu

identified after the linkage and the cases identified only on Sinan, with regard to sex, ethnicity/skin color and education level (Table 2).

Discussion

In this study, the probabilistic linkage between databases enabled the improvement of information on tuberculosis-diabetes comorbidity not registered on the national Sinan-TB database. Its results show the importance of the integration of Sinan-TB with other databases, especially with those which have information on conditions associated with tuberculosis, such as diabetes.[29] Even though Hiperdia was the system used for recording information on individuals with diabetes from 2002 to 2013 and Sinan had a field on diabetes associated with tuberculosis, there are no studies in the literature, nor publications of the Ministry of Health, that addressed the linkage between these databases.

The nominal database linkage is usually conducted by applying probabilistic methods that use pairing processes of two or more databases, with agreement probabilities – or not – from a set of common variables to those bases, searching for pairs of records related to the same individual, in addition to the standardization and qualification of information.[28]

Since the field 'condition associated with tuberculosis' is not of mandatory filling on Sinan, this record presented high percentage of blank forms. Thus, this analysis may be affected and, therefore, the lack of knowledge on comorbidity may have hampered better care, besides impairing household investigation of new cases. It is important to emphasize that recording the aforementioned field is extremely important: the increase of this comorbidity has been growing and healthcare and monitoring of cases have become more necessary each day.

Disagreement in records found between both systems shows the need to improve the integration and the flow of data among health information systems, which was also observed in studies with other official systems.[18,30]

Even with the unavailability of computers with enough memory to handle databases in this study with large amounts of records, the division of databases into six blocks enabled the linkage between Sinan-TB and Hiperdia without causing losses to the results found. However, for further studies, we recommend the use of a more efficient strategy for the databases linkage, by adopting, in every phase, a blocking key (indexing) composed of at least four fields: soundex of first name, last name, mother's first name; year of birth; sex; and municipality of residence. The use of this strategy in
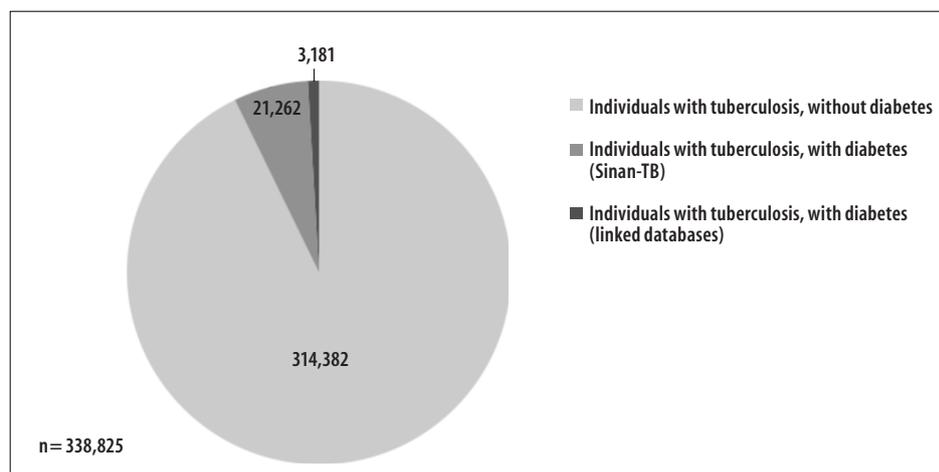
Figure 2 – Number of individuals with tuberculosis, with and without diabetes, according to the comorbidity identification database, Brazil, 2007-2011

Table 2 – Distribution of patients with tuberculosis-diabetes mellitus comorbidity according to database origin and sociodemographic variables, Brazil, 2007-2011 (n= 24.443)[a]

| Sociodemographic variables[a] | | Tuberculosis-diabetes mellitus comorbidity | | P-value[d] |
| | | Linkage Sinan-TB[b] and Hiperdia[c] | Sinan-TB[b] | |
		n (%)	n (%)	
Sex	**Male**	1,820 (57.2)	13,410 (63.1)	<0.001
	Female	1,361 (42.8)	7,850 (36.9)	
Age group (in years)	≤19	34 (1.1)	297 (1.4)	0.002
	20-39	385 (12.3)	2,918 (13.9)	
	40-59	1,643 (52.3)	11,120 (53.1)	
	≥60	1,078 (34.3)	6,593 (31.5)	
Ethnicity/ skin color	White	876 (30.3)	7,822 (42.0)	<0.001
	Black/Brown	1,978 (68.4)	10,446 (56.1)	
	Others	39 (1.3)	349 (1.9)	
Education level (in years of schooling)	≤4	428 (19.7)	1,424 (10.6)	<0.001
	5-8	1,705 (78.4)	11,713 (87.0)	
	≥8	43 (2.0)	319 (2.4)	
Regular alcohol intake	Yes	369 (13.5)	2,908 (15.1)	0.029
	No	2,356 (86.5)	16,303 (84.9)	
HIV[e]	Positive	120 (8.2)	966 (8.6)	0.583
	Negative	1,347 (91.8)	10,258 (91.4)	

a) Number of valid observations
b) Sinan-TB: Information System for Notifiable Diseases – tuberculosis
c) Hiperdia: System of Registration and Monitoring of Hypertension and Diabetes Mellitus
d) Pearson's Chi-square test
e) HIV: human immunodeficiency virus

database linkage allows the development of different keys, with at least four fields.

As limitations of this study, we should highlight the incompleteness of variables used for the linkage and the use of secondary data, and consequently, the possibility of underreporting, as we cannot say that the number of individuals with both diseases is well represented in the linked database.

Among individuals with tuberculosis and diabetes, as observed in other studies on tuberculosis, there was a predominance of the ethnicity/skin color black/brown and males; unlike diabetes in Brazil, to which the female population presents higher prevalence.[5,6] Possibly, women with diabetes registered on Hiperdia kept stronger connection with the health services, hence their greater participation in that database when comparing both sexes. The results of this study confirm the importance of linking information arising from different health information systems, especially for the detection and identification

of cases registered in other systems – not captured by the surveillance of the disease –, contributing to the improvement of SUS management.

At last, the application of database linkage can become an important tool for monitoring the health situation of individuals with tuberculosis and diabetes in treatment. Future studies should deepen the understanding of the different profiles of this comorbidity.

Authors' Contributions

RG Abreu and Sousa AIA contributed to the study design, analysis and interpretation of data, preparation of drafts and final manuscript version. Oliveira MRF and Sanchez MN contributed to the study design, interpretation of data and writing review. All authors approved the manuscript's final version and declared to be responsible for all aspects of the study, ensuring its accuracy and integrity.

References

1. Seiscento M. Tuberculose em situações especiais: HIV, diabetes mellitus e insuficiência renal. Pulmão RJ. 2012;21(1):23-6.

2. Organização Pan-Americana da Saúde. Manejo conjunto de tuberculose e diabetes mellitus: enfoque colaborativo nos serviços de saúde das cidades de Tilhuana (México) e a Bahia e São Paulo (Brasil) [Internet]. Brasília: Organização Pan-Americana da Saúde; 2012 [citado 2015 out 14]. Disponível em: http://www.paho.org/hq/index.php?option=com_docman&task=doc_view&gid=22820&Itemid=

3. Brasil. Ministério da Saúde. Portaria nº 204, de 17 de fevereiro de 2016. Define a Lista Nacional de Notificação Compulsória de doenças, agravos e eventos de saúde pública nos serviços de saúde públicos e privados em todo o território nacional, nos termos do anexo, e dá outras providências. Diário Oficial da República Federativa do Brasil, Brasília (DF), 2016 fev 18; Seção 1:23.

4. Maciel ELN, Sales CMM. A vigilância epidemiológica da tuberculose no Brasil: como é possível avançar mais? Epidemiol Serv Saude. 2016 jan-mar;25(1):175-8.

5. World Health Organization. 2015 Global tuberculosis report [Internet]. Geneva: World Health Organization; [2016] [cited 2016 Jun

28]. Available from: http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf

6. Ministério da Saúde (BR). Secretaria de Vigilância em Saúde. Departamento de Vigilância das Doenças Transmissíveis. Panorama da tuberculose no Brasil: indicadores epidemiológicos e operacionais. Brasília: Ministério da Saúde; 2014.

7. Ministério da Saúde (BR). Secretaria de Vigilância em Saúde. Detectar, tratar e curar: desafios e estratégias brasileiras frente à tuberculose. Bol Epidemiol. 2015;46(9):1-19.

8. Organização Mundial da Saúde. Cuidados inovadores para condições crônicas: componentes estruturais de ação. Brasília: Organização Mundial da Saúde; 2003.

9. Ministério da Saúde (BR). Secretaria de Atenção à Saúde. Departamento de Atenção Básica. Estratégias para o cuidado da pessoa com doença crônica: diabetes mellitus. Brasília: Ministério da Saúde; 2013. (Cadernos de Atenção Básica, 36).

10. Ministério da Saúde (BR). Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Manual de recomendações para o controle da tuberculose no Brasil. Brasília: Ministério da Saúde; 2011. (Série A. Normas e Manuais Técnicos).

11. Laguardia J, Domingues CMA, Carvalho C, Lauerman CR, Macário E, Glatt R. Sistema de Informação

de Agravos de Notificação (Sinan): desafios no desenvolvimento de um sistema de informação em saúde. Epidemiol Serv Saude. 2004 jul-set;13(3):135-46.

12. Ferreira CLRA, Ferreira MG. Características epidemiológicas de pacientes diabéticos da rede pública de saúde - análise a partir do sistema HiperDia. Arq Bras Endocrinol Metab. 2009 fev;53(1):80-6.

13. Ministério da Saúde (BR). Secretaria Executiva. Departamento de Informática do SUS. HiperDia: Sistema de Cadastramento e Acompanhamento de Hipertensos e Diabéticos: manual de operação. Rio de Janeiro: Ministério da Saúde; 2002.

14. Ministério da Saúde (BR). e-SUS Atenção Básica [Internet]. Brasília: Ministério da Saúde; 2012 [citado 2015 nov 21]. Disponível em: http://dab.saude.gov.br/portaldab/esus.php.

15. Machado JP, Silveira DP, Santos IS, Piovesan MF, Albuquerque C. Aplicação da metodologia de relacionamento probabilístico de base de dados para a identificação de óbitos em estudos epidemiológicos. Rev Bras Epidemiol. 2008 mar;11(1):43-54.

16. Coeli CM, Pinheiro RS, Camargo Júnior KR. Conquistas e desafios para o emprego das técnicas de record linkage na pesquisa e avaliação em saúde no Brasil. Epidemiol Serv Saude. 2015 out-dez;24(4):795-802.

17. Pinheiro RS, Andrade VL, Oliveira GP. Subnotificação da tuberculose no Sistema de Informação de Agravos de Notificação (Sinan): abandono primário de bacilíferos e captação de casos em outras fontes de informação usando linkage probabilístico. Cad Saude Publica. 2012 ago;28(8):1559-68.

18. Bartholomay P, Oliveira GP, Pinheiro RS, Vasconcelos AMN. Melhoria da qualidade das informações sobre tuberculose a partir do relacionamento entre bases de dados. Cad Saude Publica. 2014 nov;30(11):2459-69.

19. Pinheiro RS, Andrade VL, Oliveira GP. Subnotificação da tuberculose no Sistema de Informação de Agravos de Notificação (SINAN): abandono primário de bacilíferos e captação de casos em outras fontes de informação usando linkage probabilístico. Cad Saude Publica. 2012 ago;28(8):1559-68.

20. Ministério da Saúde (BR). Secretaria de Vigilância em Saúde. Guia de Vigilância em Saúde. Brasília: Ministério da Saúde; 2014.

21. Ministério da Saúde (BR). Sistema de Gestão Clínica de Hipertensão Arterial e Diabetes Mellitus da Atenção Básica (Hiperdia) [Internet]. Brasília: Ministério da Saúde; ANO [citado 2016 out 24] Disponível em: http://hiperdia.datasus.gov.br/.

22. Brasil. Ministério da Saúde. Portaria Conjunta nº 02/SPS/SE, de 05 de março de 2002. Disponibiliza, no âmbito da atenção básica, instrumento de cadastro e acompanhamento dos portadores de Hipertensão Arterial e Diabetes Mellitus, Sishiperdia, para utilização pelos municípios. Diário Oficial da República Federativa do Brasil, Brasília (DF), 2002 mar 20;Seção 1:38.

23. Brasil. Ministério da Saúde. Portaria Conjunta nº 112/SPS/SE, de 19 de junho de 2002. Aprova o fluxo de alimentação da base nacional do Hiperdia, obrigatório para todos os municípios que aderirem ao Programa de Assistência Farmacêutica à Hipertensão Arterial e ao Diabetes Mellitus conforme estabelecido no anexo da Portaria. Diário Oficial da República Federativa do Brasil, Brasília (DF), 2002 jun 20;Seção 1:60.

24. Coeli CM, Camargo Júnior KR. Avaliação de diferentes estratégias de blocagem no relacionamento probabilístico de registros. Rev Bras Epidemiol. 2002;5(2):185-96.

25. Lucena FFA, Fonseca MGP, Sousa AIA, Coeli CM. O relacionamento de bancos de dados na implementação da vigilância da aids. Cad Saude Publica. 2006 abr-jun;14(2):305-12.

26. Queiroz OV, Guerra Júnior AA, Machado CJ, Andrade ELG, Meira Júnior W, Acúrcio FA, et al. A construção da Base Nacional de Dados em Terapia Renal Substitutiva (TRS) centrada no indivíduo: relacionamento dos registros de óbitos pelo subsistema de Autorização de Procedimentos de Alta Complexidade (Apac/SIA/SUS) e pelo Sistema de Informações sobre Mortalidade (SIM) – Brasil, 2000-2004. Epidemiol Serv Saude. 2009 abr-jun;18(2):107-20.

27. Peres SV, Latorre MRDO, Michels FAS, Tanaka LF, Coeli CM, Almeida MF. Determinação de um ponto de corte para a identificação de pares verdadeiros pelo método probabilístico de linkage de base de dados. Cad Saude Colet. 2014 out-dez;22(4):428-36.

28. Silveira DP, Artmann E. Acurácia em métodos de relacionamento probabilístico de bases de dados em saúde: revisão sistemática. Rev Saude Publica. 2009 out;43(5):875-82.

29. Oliveira GP, Pinheiro RS, Coeli CM, Barreira D, Codenotti SB. Uso do sistema de informação sobre mortalidade para identificar subnotificação de casos de tuberculose no Brasil. Rev Bras Epidemiol. 2012 set;15(3):468-77.

30. Queiroz OV, Guerra Júnior AA, Machado CJ, Andrade EIG, Meira Júnior W, Acurcio FA, et al. Relacionamento de registros de grandes bases de dados: estimativa de parâmetros e validação dos resultados, aplicados ao relacionamento dos registros das autorizações de procedimentos ambulatoriais de alta complexidade com os registros de sistema de informações hospitalares. Cad Saude Colet. 2010;18(2):298-308.

Erratum

In the article "Tuberculosis and diabetes: probabilistic linkage of databases to study the association between both diseases", doi No. 10.5123/S1679-49742017000200013, published on *Epidemiology and Health Services*, 26(2), the paragraph below:

There was no statistically significant difference (p>0.05) between the cases of comorbidity identified after the linkage and the cases identified only on Sinan, with regard to sex, ethnicity/skin color and education level (Table 2).

Should read:

There was no statistically significant difference (p<0.05) between the cases of comorbidity identified after the linkage and the cases identified only on Sinan, with regard to serology for HVI (Table 2).

Access the original article at: http://www.scielo.br/pdf/ress/v26n2/en_2237-9622-ress-26-02-00359.pdf