http://dx.doi.org/10.5588/ijtld.19.0094

Supp data on figshare: https://figshare.com/s/6e84b894077da90f4f78

Completeness of notification of tuberculosis in Portugal 2015: an inventory and capture-recapture study

C. Carvalho,[1] S. Alba,[2] R. Harris,[3] I. Abubakar,[4] R. Van Hest,[5] A. M. Correia,[6] G. Gonçalves,[1] R. Duarte[7,8]

[1]Multidisciplinary Unit for Biomedical Research (UMIB), Institute of Biomedical Sciences Abel Salazar (ICBAS), University of Porto, Porto, Portugal; [2]Royal Tropical Institute, KIT Health, Amsterdam, The Netherlands; [3]National Infection Service, Public Health England, London, [4]Institute for Global Health, University College of London, London, UK; [5]Department of Tuberculosis Control, Regional Public Health Service (GGD) Groningen, Groningen, The Netherlands; [6]Braga Health Centre Group, Portuguese Northern Regional Health Administration Cávado I, Braga, [7]EPIUnit, Institute of Public Health, University of Porto (ISPUP), Porto, [8]Public Health Science and Medical Education Department, Faculty of Medicine, University of Porto, Porto, Portugal

Correspondence to: Carlos Carvalho, Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto, Departamento de Estudo das Populações, Rua Jorge de Viterbo Ferreira 228, 4050-313 Porto, Portugal. email: cfcarvalho@icbas.up.pt

Running head: Completeness of TB notification in Portugal

Article submitted 20 February 2020. Final version accepted 27 April 2020.

SUMMARY

BACKGROUND: Despite the steady decline in the last few decades, Portugal remains the Western European country with the highest TB notification rates. The aim of this study was to estimate the completeness of notification to the National Tuberculosis Programme (NTP) Surveillance System (SVIG-TB) in 2015.

METHODS: We implemented an inventory study and a three-source log-linear capture-recapture analysis using two additional data sources that were deterministic and probabilistically lined: the national notifiable diseases surveillance system (*Sistema Nacional de Vigilância Epidemiológica*, SINAVE) and the national hospital discharge database (*Grupos de Diagnósticos Homogéneos*, GDH).

RESULTS: We identified 2328 unique probable/confirmed TB cases across the three data soources. We found a positive dependency between SVIG-TB and SINAVE (incidence rate ratio [IRR] 8.9, 95% CI 6.6–12.0) and between GDH and SINAVE (IRR 2.6, 95% CI 2.0–3.4). After adjusting for these dependencies, we estimated that 266 cases (95% CI 198–358) were not reported, indicating a notification (to SVIG-TB) completeness rate of 77%.

CONCLUSION: True incidence rate of TB in Portugal in 2015 could have been as high as 26.1 per 100,000. This could be an overestimation because of false-positive cases recorded in both SINAVE and GDH or due to a smaller scale, false non-matches. Studies aimed at validating potentially false-positive cases should be implemented to address these limitations.

KEY WORDS: TB; completeness of notification; record-linkage; capture-recapture; Portugal

Notification rates form the basis of TB incidence estimates, which are the primary indicator to monitor the new global strategy and targets for TB prevention, care and control after 2015 (the End TB Strategy).[1] In Portugal, incidence rates of TB have been estimated and internationally reported by the National TB Programme (NTP) using as single data source the NTP surveillance system SVIG-TB (*Sistema de Vigilância da Tuberculose*).[2] TB notification rates have been steadily declining in the last decades in Portugal, reaching 20.1 notifications per 100,000 population in 2015 (close to the cut-off between intermediate- and low-incidence in a region).[3]

According to the 2017 European Centre for Disease Control and Prevention Surveillance Report, however, the true incidence rate in Portugal would be 15% higher (23/100,000 in 2015) due to under-notification.[4] When multiple data sources are available, it is possible to estimate the completeness of notification using inventory studies, combined or not with capture-recapture (CRC) modelling.[5,6]

In the present study, we aimed at assessing the completeness of notification to the NTP Surveillance System and estimating the true incidence rate of TB in Portugal in 2015, implementing an inventory study and a CRC analysis with two additional national TB data sources.

METHODS

Design and study population

We conducted an inventory study and CRC analysis including all records of diagnosed TB, retrieved from three national electronic, case-based databases for 2015 in mainland Portugal. Autonomous Island Regions of Azores and Madeira were excluded as there were no data on hospital discharges in these regions.

Case definition

We included all notified cases of active TB in mainland Portugal in 2015. Cases were considered 1) confirmed using a positive culture for *Mycobacterium tuberculosis* complex (gold standard) or polymerase chain reaction (PCR) and sputum smear; 2) probable, if clinical and laboratory results (PCR, microscopy or histopathological findings) were compatible with TB; or 3) possible, if diagnosis was made on clinical, radiological and/or epidemiological arguments.

Data sources

NTP surveillance system (SVIG-TB)

This system is based on the clinical notification by physicians from TB outpatient centres, which are part of the Portuguese National Health Service. Cases are notified and monitored using two different paper forms and sent to "notification centres" for computer recording and subsequent national data aggregation. At the national level, all data are anonymised, but information on date of birth, sex and place of residence are kept (as well as clinical information).

Patients with suspect TB are referred to TB outpatient centres from hospitals, general practitioners, private practitioners, public health services or by their own initiative. If suspicion is high enough treatment is started, triggering a first notification to SVIG-TB. When culture and drug susceptibility testing (DST) results are available, a second notification to SVIG-TB is triggered. There is a continuous update with serious adverse events, laboratory results and outcome (Figure 1). Any false diagnostic of TB (non-tuberculous mycobacterial infection or other diagnosis) is recorded in "treatment outcome" and the case is excluded ("de-notified").

National notifiable diseases surveillance system (SINAVE)

TB is a statutory notifiable disease in Portugal since 1902.[7] In 2014 a new web-based notification platform, SINAVE (*Sistema Nacional de Vigilância Epidemiológica*), was introduced.[8] All physicians from the public and private sectors are legally required to notify to SINAVE whenever they suspect TB. The primary purpose of this surveillance system is triggering epidemiological investigation, risk assessment and community intervention by public health services. In theory, SINAVE should be updated with all available information before cases are validated as possible, probable, confirmed or non-cases (Figure 2). However, as culture results usually take some weeks, some cases are validated before there is enough information to accurately classify or exclude TB diagnosis.

*Hospital discharge database (*Grupos de Diagnósticos Homogéneos*)*

This database is managed by the Portuguese Central Administration of the Health System. All diagnoses made in the Portuguese National Health Service hospitals are coded upon discharge, using the International Statistical Classification of Diseases, Injuries and Causes of Death (ICD) version 9 before 2016 and version 10 since then. ICD9 codes referring to diagnoses of TB start with 010–018 (TB), 37105 (Phthisical cornea) or 6473 (TB in pregnancy).

Record linkage

All records of TB diagnosed in the study period were received in electronic spreadsheet format, including the variables date of birth, sex, place of residence (district, municipality and parish),

date of diagnosis (estimated in *Grupos de Diagnósticos Homogéneos* [GDH], the hospital discharge database, as date of admission plus one third of hospitalisation days), site of disease and laboratory results (ICD9 codes termination digits in GDH), HIV co-infection (ICD9 codes started by "042", "07953" or "V08" in *GDH*) and vital status at discharge.

Data were abstracted from each data source for cases diagnosed from 1 July 2014 to 30 June 2016 to allow for correction of late notification. After record linkage, records from 1 July 2014 to 31 December 2014 and from 1 January to 30 June 2016 found in only one data source were excluded.

An initial step of deterministic linkage was performed,[6] matching exactly date of birth, sex and place of residence. These variables were expected to enable unique identification of a single case across the three databases.

To account for possible recording errors in the matching variables (records that in reality referred to the same case, although there were actually differences in the data recorded), probabilistic matching was performed using the Stata (StataCorp, College Station, TX, USA) user-written command *reclink*.[9] Agreement or disagreement weights were assigned to specific properties of each record (to accommodate different probability of error in different fields), which were then combined into a single probability score. Linked records were visually inspected, sorted by calculated probability scores, and different matching variables and minimum matching weights were used until most links were considered acceptable.

As a first step, the GDH data set was merged with SVIG-TB and thereafter the resulting dataset was merged with the SINAVE data set (minimum overall matching score of 0.87 to declare a match). The following relative weights were used: first digit of day of birth (10); second digit of day of birth (10); first digit of month of birth (10); second digit of month of birth (10); third digit of year of birth (15); fourth digit of year of birth (10); sex (15); district of residence (15); municipality of residence (5); parish of residence (3).

Cases were allocated to one of seven different strata (case existing in data sets A, B and C, A and B, A and C, B and C, only A, only B and only C) and assigned an unique case-classification depending on the information available: confirmed, if confirmed in at least one data source; probable, if not confirmed but probable in at least one data source; possible, if not confirmed nor probable in any of the three data sources.

Capture-recapture analysis

CRC analyses rely on Poisson regression models to estimate the total number of unreported cases in a data system using the information from all available data sources. The data consist

of aggregated totals for each data set and overlaps between data sets (i.e., all different portions of the Venn Diagram in Figure 3). The models are built in a way that accounts for the fact that some data sets are "dependent" on each other, meaning that the probability of a case being recorded in one data set is related to the probability of being recorded in another data set. Such dependencies can result in biased estimates of the number of unreported cases if not accounted for. The incidence rate ratio (IRR) of the interaction term for belonging to two data sets is thus a measure of dependency between databases: if the IRR is greater than 1, the conclusion is that the two sources are positively associated.[10,11]

The models fitted in this study included three potential dependencies between pairs of datasets. As a result, eight possible models were considered, ranging from no dependencies (the base model) to all three (saturated model). The choice of model was determined by balancing model fit with parsimony, based on Akaike Information Criterion (AIC) scores.[12] The chosen model was further expanded and analyses stratified by the covariates sex, region of residence, case classification, site of disease, HIV coinfection and vital status in turn.

Capture-recapture analysis was repeated after excluding possible cases. This was done in order to limit the impact of including potentially false-positive cases in the analysis.[13–15] All data management, including record linkage and CRC analysis, was performed using Stata v15.[16]

Ethics approval

This study received clearance by the Portuguese National Data Protection Committee (no 1043/2017) and was approved by the Northern Regional Health Administration Ethical Board (no 114/2017).

RESULTS

From 1 January to 31 December 2015, 2086 cases were diagnosed with TB and subsequently notified to the National Tuberculosis Programme (NTP) surveillance system (SVIG-TB). In that same period, 1874 cases were recorded in the national notifiable diseases surveillance system (SINAVE) and 1358 in the hospital discharge database (GDH) (Table 1).

Identification of matches

Using deterministic record-linkage, 3561 unique records were identified across all three datasets. Most of the unmatched cases were observed in the hospital discharge database (61.9% of cases admitted to hospital) (Supplementary Table S1). The probabilistic record-linkage between the three data sets (matching scores shown on Table 2) greatly increased the overlap

between data sources, allowing the identification of 2786 unique records (Supplementary Table S2). The pattern of overlap between data sources varied across regions and case classifications (Supplementary Tables S3 and S4). Considering only probable and confirmed cases, 2328 unique records were identified (Figure 3).

Capture-recapture analysis

The model with better fit included two-way interactions between SVIG-TB and SINAVE, as well as between GDH and SINAVE (Table 3). Both indicated a positive dependence between sources (probabilistically linked data set, including only probable and confirmed cases—SVIG-TB:SINAVE, IRR 8.9 (95%CI 6.6–12.0); SINAVE:GDH IRR 2.6 (95%CI 2.0–3.4).

The estimated number of unreported cases differed substantially across different models —but was comparable in models with similar AIC statistics. The model with all three two-way interactions (M8) had similar AIC to the chosen model (M6), and a similar number of unreported cases. This suggests that M6 and M8 have similar performance and that the choice of a more parsimonious model does not affect accuracy (Table 3).

The inclusion of covariates in the model made little difference in the overall estimated number of unreported cases, but those covariates affected capture probabilities and overlap patterns. Unreported cases were more likely to be possible cases (as opposed to being confirmed or probable), HIV-negative and alive (Table 4).

Completeness of notification (sensitivity)

Taking into account the figures above, the estimated sensitivity of SVIG-TB for probable/confirmed cases of TB was 77.0% [1997/(2328+266)] (Supplementary Table S5).

DISCUSSION

This study was a first attempt at evaluating the sensitivity of the surveillance systems for tuberculosis in Portugal. It included a comprehensive CRC analysis using log-linear regression, based on three data sources both deterministically and probabilistically linked using two different case definitions. We included in the analysis various covariates, which showed different capture probabilities and between-source dependencies but had little impact on the overall estimate of unreported cases.

The high dependency between SVIG-TB and SINAVE is consistent with the process of notification, since all physicians are legally required to notify new TB cases to Public Health

(through SINAVE). The positive dependency between GDH and SINAVE may also imply that hospitalised cases are more likely to be notified to SINAVE (rather than the reverse).

Probable errors in data entry to the data sources prevented accurate deterministic data linkage and hence the capture-recapture analysis in this data set produced an unrealistically high estimate of the number of unreported TB cases. This is an important limitation of CRC methods[17] that we addressed using probabilistic record-linkage. Applying the CRC analysis to a probabilistically merged dataset led to the estimation of 781 unreported cases of TB (95%CI 618–986). This seemed to be still an overestimation of unreported cases and could be a result of including potentially false-positive cases in the analysis, as observed in other studies.[13–15]

An important assumption for application of CRC methods is that all records should correspond to true cases.[17–20] We suspect that at least some cases observed in our study, especially those from the GDH, are false-positive—possibly non-tuberculous mycobacterial infections and/or diagnostic miscoding.[14,15] Unreported cases were more likely to have been possible cases and that most cases observed only in GDH were classified as possible, which supports that hypothesis.

To address this limitation and produce a more reliable estimate of the completeness of notification, we applied the CRC analysis to a data set, including only cases that could be classified as probable or confirmed. Assuming this more specific case definition, completeness of notification to SVIG-TB in Portugal would be 77.0%, corresponding to an incidence rate of 26.1 TB cases/100,000 population in 2015 (30% above what was reported to WHO and ECDC, which was 20.1/100 000 population).

Although this is a more acceptable estimate, it could still not correspond to the truth. In a CRC study, all cases should have equal probability of being captured by any of the data sources,[19,21–24] which is not true for the GDH (as only the most severe cases of TB are admitted to hospital).[25] Furthermore, there should not be subgroups with very different probabilities of being observed in one data source and re-observed in another data source.[21,26,27] In this respect, it is reassuring that the estimated number of unreported cases did not differ after the inclusion of a number of covariates that could be potential sources of heterogeneity.

Our study brings important considerations for TB surveillance in Portugal. CRC methods provide feasible way of correcting incidence rates for under-notification,[28,29] but this should not replace the effort to improve the sensitivity of the NTP surveillance system and continuously monitor the quality of the data. CRC analysis would produce the best estimates of the true incidence of TB and could be the recommended method to produce national statistics if data sources were 100% specific. Regarding sensitivity, previous studies suggested possible

reasons for under-notification of TB.[30] This was not addressed in our study but will be a research priority.

From an operational research perspective, another study should be conducted to clarify "potential false positive cases", to make sure that they were true cases of tuberculosis and their matching variables were recorded correctly. In the meantime, as SVIG-TB combines a relatively high sensitivity with a higher specificity, it should remain as the main data source used for TB surveillance.

Acknowledgements

The authors thank M Gomes and C Sousa Pinto for their contributions extracting and providing data from the national TB registries; and M Bakker, C Mergenthaler, E Rood and M Straetemans for their scientific support to this project.

This work was supported by the European Centre for Disease Prevention and Control (ECDC/2016/004—Framework Contract "Assessment of tuberculosis underreporting through inventory studies").

Conflicts of interest: none declared.

References

1 Uplekar M, Weil D, Lönnroth K, et al. WHO's new end TB strategy. Lancet 2015; 385(9979): 1799–1801.

2 Portuguese Directorate-General of Health. Tuberculose: Sistema SVIG-TB. Lisbon Portugal: Portuguese Directorate-General of Health, https://www.dgs.pt/paginas-de-sistema/saude-de-a-a-z/tuberculose1/sistema-svig-tb.aspx. Accessed for the last time today (5Sep2020)

3 Portuguese Directorate-General of Health. Tuberculose em Portugal: desafios e estratégias, 2018. Lisbon Portugal: Portuguese Directorate-General of Health, 2018. https://www.dgs.pt/documentos-e-publicacoes/tuberculose-em-portugal-desafios-e-estrategias-2018-.aspx.

4 European Centre for Disease Prevention and Control, WHO Regional Office for Europe. Tuberculosis surveillance and monitoring in Europe, 2017. Stockholm, Sweden: ECDC, 2017.

5 van Hest R. Capture-recapture methods in surveillance of tuberculosis and other infectious diseases. Rotterdam: Erasmus MC, Univ Med Cent Rotterdam Repos, 2007.

6 World Health Organization. Assessing tuberculosis under-reporting through inventory studies. Geneva, Switzerland: WHO, 2012.

7 Portugal Inspeção Geral dos Serviços Sanitários do Reino. Regulamento geral dos serviços de saúde e beneficiência pública. Lisbon, Portugal: 1902: pp 3–126.

8 Portuguese Directorate-General of Health. Sistema Nacional de Vigilância Epidemiológica (SINAVE). Lisbon Portugal: Portuguese Directorate-General of Health, https://www.dgs.pt/servicos-on-line1/sinave-sistema-nacional-de-vigilancia-epidemiologica.aspx.

9 Blasnik M. RECLINK: Stata module to probabilistically match records. Stat Softw Components S456876, 2010. Boston: Boston College Department of Economics

10 Cormack RM. Log-linear models for capture-recapture. Biometrics 1989; 45(2): 395-413.

11 Fienberg SE. The multiple recapture census for closed populations and incomplete 2 contingency tables. Biometrika 1972; 59(3): 591-603.

12 Hook EB, Regal RR. Validity of methods for model selection, weighting for model uncertainty, and small sample adjustment in capture-recapture estimation. Am J Epidemiol 1997; 145(12): 1138–1144.

13 Tocque K, Bellis MA, Beeching NJ, Davies PD. Capture recapture as a method of determining the completeness of tuberculosis notifications. Commun Dis Public Health 2001; 4(2):141-143.

14 Van Hest NAH, Smit F, Baars HWM, et al. Completeness of notification of tuberculosis in The Netherlands: how reliable is record-linkage and capture–recapture analysis? Epidemiol Infect 2007; 135(6): 1021–1029.

15 Van Hest NAH, Story A, Grant AD, Antoine D, Crofts JP, Watson JM. Record-linkage and capture-recapture analysis to estimate the incidence and completeness of reporting of tuberculosis in England 1999–2002. Epidemiol Infect 2008; 136(12): 1606-1616.

16 Statacorp. Stata Statistical Software: Release 15. College Station, TX, USA: StataCorp, 2017.

17 Desenclos J-C, Hubert B. Limitations to the universal use of capture-recapture methods. Int J Epidemiol 1994; 23(6): 1322–1323.

18 Dunn J, Andreoli SB. Método de captura e recaptura: nova metodologia para pesquisas epidemiológicas. Rev Saude Publica 1994; 28(6): 449–453.

19 Cormack R. Problems with using capture-recapture in epidemiology: an example of a measles epidemic. J Clin Epidemiol 1999; 52(10): 909–914.

20 Ding Y, Fienberg SS. Multiple sample estimation of population and census undercount in the presence of matching errors. Surv Methodol 1996; 22(1): 55–64.

21 Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. Epidemiol Rev 1995; 17(2): 243–264.

22 Hook EB, Regal RR. Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. J Clin Epidemiol 1999; 52(10): 917–926.

23 Hook EB, Regal RRA. Effect of variation in probability of ascertainment by sources ("variable catchability") upon "capture-recapture" estimates of prevalence. Am J Epidemiol 1993; 137(10): 1148-1166.

24 Tilling K. Capture-recapture methods-useful or misleading? International Journal of Epidemiology 2001; 30(1): 12-14.

25 Galego MA, Santos JV, Viana J, Freitas A, Duarte R. To be or not to be hospitalised with tuberculosis in Portugal. Int J Tuberc Lung Dis 2019; 23(9): 1029–1034.

26 Chao A, Tsay PK, Lin S-HH, Shau W-YY, Chao D-YY. The applications of capture-recapture models to epidemiological data. Stat Med 2005; 1(20): 31–65.

27 Tilling K, Sterne JAC. Capture-recapture models including covariate effects. Am J Epidemiol 1999; 149(4):392-400

28 Wolter KM. Accounting for America's uncounted and miscounted. Science 1991; 253(5015): 12–15.

29 Dunn J, Andreoli SB. Capture and recapture method: a new methodology for epidemiological research. Rev Saude Publica 1994; 28(6): 449–453.

30 Li T, Shewade HD, Soe KT, et al. Under-reporting of diagnosed tuberculosis to the national surveillance system in China: an inventory study in nine counties in 2015. BMJ Open 2019; 9:e021529

Table 1 Cases of TB notified to SVIG-TB and SINAVE and admitted to hospital in mainland Portugal, 1 January–31 December 2015

Variable	Categories	Cases in each of the data sources (proportion within data source) SVIG-TB *n* (%)	SINAVE *n* (%)	GDH *n* (%)
Region of residence	Northern Region (population 3.6M)	865 (41.5)	805 (43)	445 (32.8)
	Lisbon and Tagus Valley (population 3.6M)	872 (41.8)	750 (40.0)	612 (45.1)
	Central Region (population 1.7M)	179 (8.6)	175 (9.3)	181 (13.3)
	Alentejo (population 0.48M)	79 (3.8)	59 (3.1)	52 (3.8)
	Algarve (population 0.44M)	91 (4.4)	85 (4.5)	68 (5.0)
Sex	Male	1391 (66.7)	1246 (66.5)	940 (69.2)
	Female	695 (33.3)	628 (33.5)	418 (30.8)
Age group, years	0–5	11 (0.5)	9 (0.5)	17 (1.3)
	5–15	21 (1.0)	21 (1.1)	19 (1.4)
	15–25	182 (8.7)	154 (8.2)	105 (7.7)
	25–35	248 (11.9)	217 (11.6)	124 (9.1)
	35–45	409 (19.6)	374 (20)	247 (18.2)
	45–55	409 (19.6)	383 (20.4)	257 (18.9)
	55–65	325 (15.6)	287 (15.3)	220 (16.2)
	65–75	214 (10.3)	183 (9.8)	148 (10.9)
	≥75	267 (12.8)	246 (13.1)	221 (16.3)
Site of disease	Pulmonary	1475 (70.7)	1301 (69.4)	881 (64.9)
	Extra-pulmonary	602 (28.9)	507 (27.1)	477 (35.1)
	Unknown	9 (0.4)	66 (3.5)	-
Case classification	Confirmed	1278 (61.3)	1222 (65.2)	147 (10.8)
	Probable	466 (22.3)	447 (23.9)	784 (57.7)
	Possible	342 (16.4)	205 (10.9)	427 (31.4)
HIV co-infection	Yes	223 (10.7)	184 (9.8)	214 (15.8)
	No/not recorded	1863 (89.3)	1690 (90.2)	1144 (84.2)
Vital status*	Dead	153 (7.3)	94 (5)	152 (11.2)
	Alive/not recorded	1933 (92.7)	1780 (95)	1206 (88.8)
Total		2086 (100)	1874 (100)	1358 (100)

* Last updated at the end of treatment (SVIG-TB), recorded by Public Health services within 1–3 months after diagnosis/notification but might be updated by the regional or national level at a later stage (SINAVE) at the date of hospital discharge (GDH).

TB = tuberculosis; SVIG-TB = *Sistema de Vigilância da Tuberculose*; SINAVE = *Sistema Nacional de Vigilância Epidemiológica*; GDH = *Grupos de Diagnósticos Homogéneos*; HIV = human immunodeficiency virus.

Table 2 Accepted matching scores (probabilistic record-linkage of tuberculosis datasets), frequency and description of corresponding mismatch between linking variables, mainland Portugal, 2015

Overall matching score	Proportion of cases matched (first record linkage, GDH with SVIG-TB)* %	Proportion of cases matched (second record linkage, GDH_SVIG-TB with SINAVE)† %	Mismatch between linking variables‡
1.0000	15.7	42.0	Perfect match
0.9709	18.7	20.6	Different parish
0.9223	1.5	2.7	Different municipality (and parish)
0.9029	1.1	1.4	One digit of date of birth
0.8738	2.6	1.9	One digit of date of birth and parish of residence
<0.8700	60.5	31.4	Rejected match

*Denominator includes all non-matched records from SVIG-TB and GDH plus matched records (*n* = 2668).

†Denominator includes all non-matched records from the database generated in Column 1 (SVIG-TB_GDH) and SINAVE plus matched records (*n* = 2786).

‡Variables used for record-linkage: date of birth (day, month, year), sex and place of residence (district, municipality, parish).

GDH = *Grupos de Diagnósticos Homogéneos*; SVIG-TB = *Sistema de Vigilância da Tuberculose*; SINAVE = *Sistema Nacional de Vigilância Epidemiológica*.

Table 3 Model selection statistics and estimated number of unreported cases of TB in the probabilistically record-linked dataset, mainland Portugal, 2015

		Possible, probable and confirmed cases			Probable/confirmed cases		
Model (interaction terms)	df	AIC	Unreported cases*	95% CI	AIC	Unreported cases*	95% CI
M1 (No interactions)	4	653.3	92.2	81.9–103.8	263.2	36.6	31.4–42.5
M2 (SVIG-TB:SINAVE)	5	150.8	333.6	291.0–382.5	121.9	116.6	95.5–142.5
M3 (SVIG-TB:GDH)	5	549.4	46.7	38.3–57.0	236.9	24.7	19.8–30.9
M4 (GDH:SINAVE)	5	654.9	94.6	81.9–109.3	252.8	43.5	36.6–51.8
M5 (SVIG-TB:SINAVE; SVIG-TB:GDH)	6	149.9	272.4	207.5–357.5	121.9	97.2	70.3–134.4
M6 (SVIG-TB:SINAVE; GDH:SINAVE)	6	64.5	780.8	618.0–986.4	62.6	266.5	198.2–358.3
M7 (SVIG-TB:GDH; GDH:SINAVE)	6	550.6	44.8	35.9–55.9	229.8	29.2	22.8–37.3
M8 (SVIG-TB:SINAVE; SVIG-TB:GDH; GDH:SINAVE)	7	66.5	756.3	531.3–1076.7	64.5	253.6	168.1–382.5

*Estimated number of unreported cases as exponentiated intercept coefficient.

TB = tuberculosis; df = degrees of freedom used by the model (i.e., number of parameters); AIC = Akaike Information Criteria; CI = confidence interval; SVIG-TB = *Sistema de Vigilância da Tuberculose*; SINAVE = *Sistema Nacional de Vigilância Epidemiológica*; GDH = *Grupos de Diagnósticos Homogéneos.*

Table 4 Estimated number of unreported cases of TB by selected covariates in the probabilistic record-linked data set, mainland Portugal, 2015

Covariates	Unreported cases* (95% CI) Possible, probable and confirmed cases	Probable/confirmed cases
None	780.8 (618.0–986.4)	266.5 (198.2–358.3)
Sex		
Male	410.5 (309.3–544.7)	149.4 (103.8–215.0)
Female	420.5 (274.5–644.0)	121.6 (71.8–206.0)
Geographical area		
Northern region	381.5 (239.2–608.5)	93.9 (50.8–173.5)
Lisbon and Tagus Valley	243.6 (173.8–341.5)	97.4 (63.6–149.0)
Rest of the country	222.9 (139.1–357.2)	82.9 (46.8–146.7)
Case classification		
Confirmed	52.5 (31.2–88.2)	52.5 (31.2–88.2)
Probable	187.6 (121.5–289.6)	187.6 (121.5–289.6)
Possible	668.5 (422.7–1057.2)	—
Site of disease		
Pulmonary	476.1 (349.2–649.0)	172.6 (118.8–250.9)
Extra-pulmonary	288.3 (201.2–413.1)	84.7 (51.6–139.2)
HIV co-infection		
HIV	52.6 (29.1–94.8)	22.3 (10.7–46.6)
No HIV	765.6 (591.9–990.3)	252.4 (182.3–349.4)
Vital status		
Dead	23.5 (11.0–50.0)	9.6 (4.0–22.6)
Alive	783.5 (608.4–1008.9)	267.0 (192.7–369.9)

*Estimated number of unreported cases as exponentiated intercept coefficient, according to Model M6.

TB = tuberculosis; CI = confidence interval; HIV = human immunodeficiency virus.

FIGURE LEGENDS

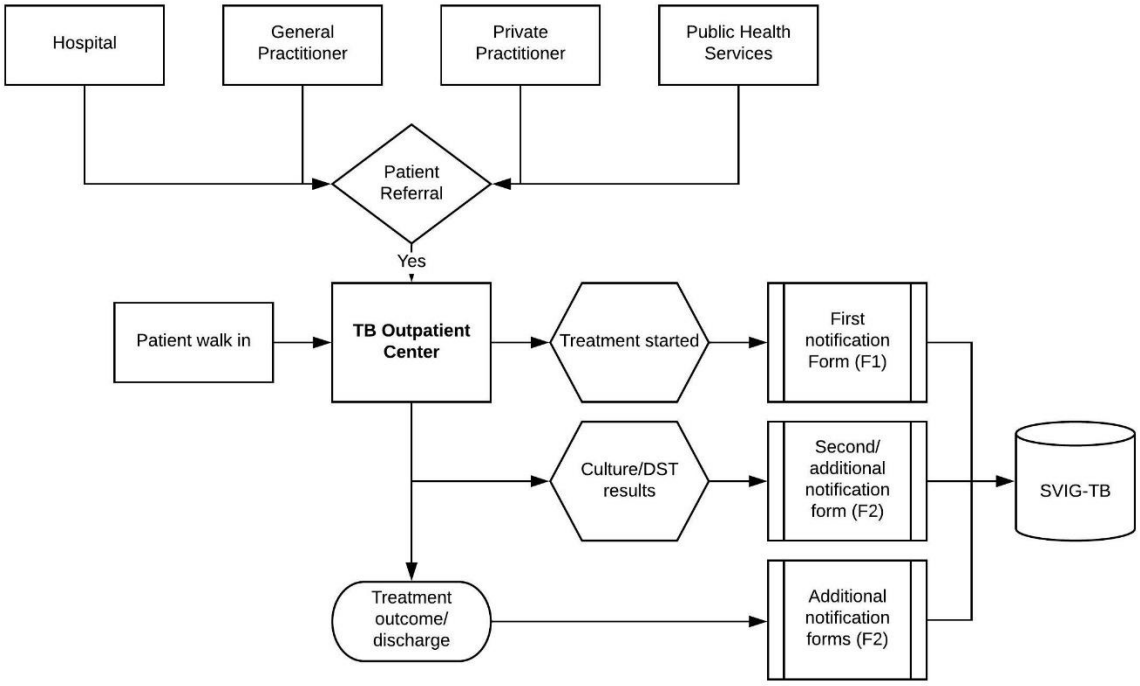

Figure 1 National Tuberculosis Programme Surveillance System (SVIG-TB) patient referral and data flow, Portugal. TB = tuberculosis; DST = drug susceptibility testing; SVIG-TB = *Sistema de Vigilância da Tuberculose*.

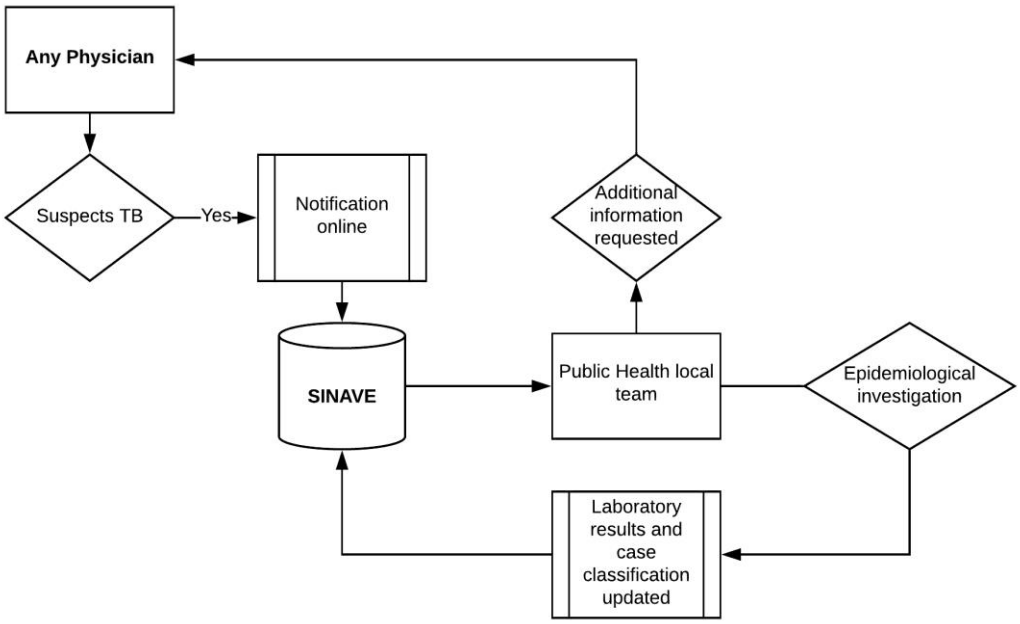

Figure 2 SINAVE data flow, Portugal. TB = tuberculosis; SINAVE = *Sistema Nacional de Vigilância Epidemiológica.*

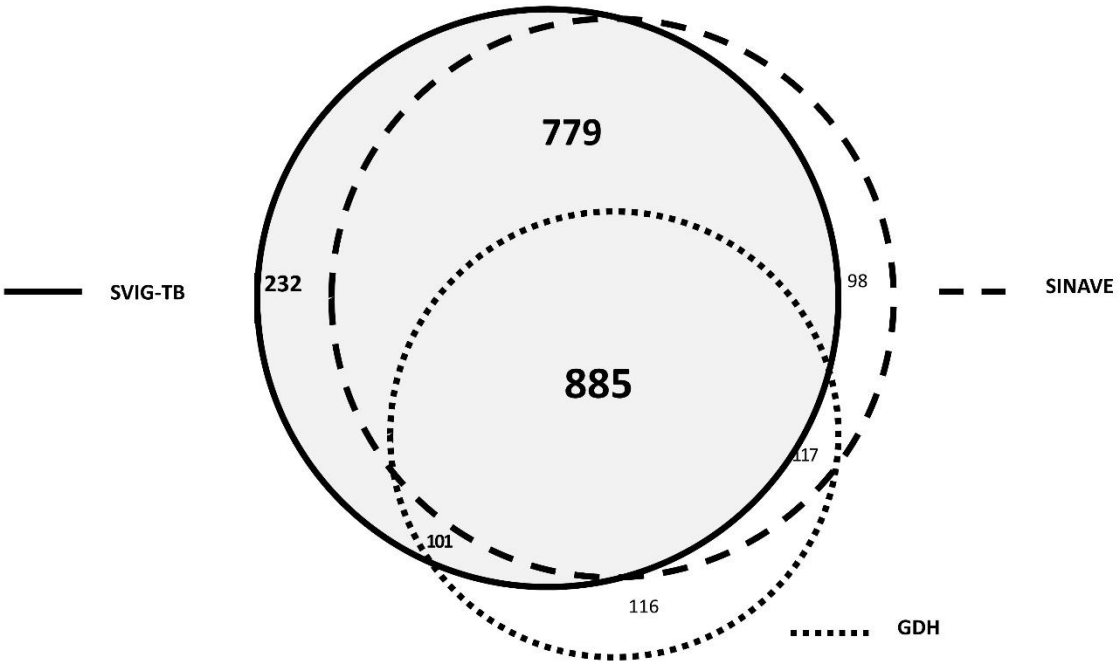

Figure 3 Probabilistic linkage of probable and confirmed cases of TB between three data sources: SVIG-TB, SINAVE and GDH, Mainland Portugal, 2015. Numbers represent the number of cases observed in each one, two or three data sources (inventory study, *n* = 2328). SVIG-TB = *Sistema de Vigilância da Tuberculose*; SINAVE = *Sistema Nacional de Vigilância Epidemiológica*; GDH = *Grupos de Diagnósticos Homogéneos.*

Figures 1 and 2 were created online on LucidChart and can be accessed via this link:
https://app.lucidchart.com/invitations/accept/79c83024-40cf-464b-97c5-80036f8c55a3
I shared the Project with iroy@theunion.org

RÉSUMÉ

CONTEXTE : En dépit d'un déclin régulier au cours des dernières décennies, le Portugal reste le pays d'Europe de l'Ouest qui a le taux le plus élevé de notification de la TB. Le but de cette étude a été d'estimer l'exhaustivité de la notification au système de surveillance du Programme National Tuberculose (NTP) (SVIG-TB) en 2015.

MÉTHODE : Nous avons mis en œuvre une étude d'inventaire et une analyse log-linéaire de capture-recapture à partir de trois sources en utilisant deux sources supplémentaires de données qui ont été liées de manière déterministe et probabiliste : le système national de surveillance des maladies à déclaration obligatoire (SINAVE) et la base de données nationale de sortie des hôpitaux (GDH).

RÉSULTATS : Nous avons identifié 2328 cas de TB unique probables/confirmés dans trois sources de données. Nous avons trouvé une dépendance positive entre SVIG-TB et SINAVE (taux d'incidence [IRR] 8,9 ; IC 95% 6,6–12,0) et entre GDH et SINAVE (IRR 2,6 ; IC 95% 2,0–3,4). Après ajustement sur ces dépendances, nous avons estimé que 266 cas (IC 95% 198–358) n'avaient pas été rapportés, correspondant à 77% d'exhaustivité de la notification à SVIG-TB.

CONCLUSION : Le taux d'incidence de la TB au Portugal en 2015 pourrait avoir atteint 26,1 pour 100 000. Ceci pourrait être une sur estimation, à cause des faux positifs enregistrés à la fois dans SINAVE et GDH ou à une plus petite échelle à cause de faux non appariés. Des études visant à valider les faux positifs potentiels devraient être mis en œuvre pour combattre ces limites.

RESUMEN

MARCO DE REFERENCIA: Pese a la disminución constante en los últimos decenios, Portugal sigue siendo el país con la tasa de notificación de casos de TB más alta en Europa occidental. El objetivo del presente estudio fue evaluar la exhaustividad del sistema de vigilancia (SVIG-TB) del Programa Nacional de Tuberculosis (PNT) en el 2015.

MÉTODO: Se llevó a cabo un estudio de inventario y un análisis de captura y recaptura de tres fuentes de información con un modelo logarítmico lineal y se usaron dos fuentes de datos adicionales con vínculo determinista y probabilístico: el sistema nacional de vigilancia de enfermedades de declaración obligatoria (SINAVE) y la base de datos nacional de altas hospitalarias (GDH).

RESULTADOS: Se encontraron 2328 casos únicos de TB probable o confirmada en las tres fuentes de datos. Se observó una dependencia positiva entre el SVIG-TB y el SINAVE (cociente de tasas de incidencia [IRR] 8,9; IC 95% 6,6-12,0) y entre la GDH y el SINAVE (IRR 2,6; IC 95% 2,0-3,4). Tras ajustar con respecto a estas dependencias, se estimó que no se habían notificado 266 casos (IC 95% 198-358), lo cual equivale a una exhaustividad de 77% de la notificación al SVIG-TB.

CONCLUSIÓN: La incidencia real de TB en Portugal en el 2015 pudo haber sido hasta de 26,1 por 100 000 habitantes. Esta cifra podría ser una sobreestimación, debido a los casos positivos falsos registrados tanto en el SINAVE como en la GDH o, en menor escala, a los falsos negativos. Con el fin de superar estas limitaciones, es necesario emprender estudios que validen los posibles casos positivos falsos.