Identifying cases of type 2 diabetes from heterogeneous data sources: strategy from the EMIF Project

G. Roberto (ARS), I. Leal (EMC), N. Sattar (BHF), K. Loomis (PFIZER), P.Avillach (EMC), P. Egger (GSK), R. van Wijngaarden (PHARMO), D. Ansell (THIN), S. Reisberg (UTARTU), M. Tammesoo (UTARTU) H. Alavere (UTARTU), A. Pasqua (GENOMEDICS), L. Pedersen (AUH), J. Cunningham (UNIMAN), L. Tramontan (PEDIANET), MA. Mayer (UPF), R. Herings (PHARMO), P. Coloma (EMC), F. Lapi (ARS), M. Sturkenboom (EMC), J. van der Lei (EMC), M. Schuemie (JANSSEN), P. Rijnbeek(EMC) and R. Gini (EMC)

Agenzia Regionale di sanità della Toscana(ARS), Erasmus University Medical Center (EMC), British Heart Foundation (BHF), Pfizer Worldwide Research and Development (Pfizer),GlaxoSmithKline,(GSK), Institute for Drug Outcomes Research (PHARMO), The Health Improvement Network (THIN), University of Tartu (UTARTU), Genomedics (GENOMEDICS), Aarhus University Hospital (AUH), University of Manchester (UNIMAN), (PEDIANET), Universitat Pompeu Fabra (UPF), Janssen Research & Development (Janssen)

Background

The European Medical Information Framework (EMIF) project is establishing an EU wide information communication technology infrastructure (EMIF-Platform) to facilitate the execution of high quality multi-data base observational studies leveraging the combination of data sources with heterogeneous characteristics, such as different database structure, contents, reasons for recording, language, coding terminologies and healthcare system organization. For this purpose, a template data derivation process was specifically developed and the identification of Type 2 diabetes mellitus (T2DM) was used as a test case.

Objectives

1) To establish a set of standard algorithms useful to identify patients with type 2 diabetes (T2DM) across heterogeneous data sources, 2) to describe the data source-tailored combinations of standard algorithms recommended by local experts, and 3) to assess the impact of each standard algorithm on the characteristics of identified cases, across different data sources.

Materials and methods

Data sources

Eight data sources from six different EU countries were included: three primary care data sources (PCDs) from Italy (PCD-I) Netherlands (PCD-N) and United Kingdom (PCD-UK) respectively, three record linkage data networks (RLDs) from Italy (RLD-I), Netherlands (RLD-N) and Denmark (RLD-DK), one hospital data source (HD) from Spain and one biobank from Estonia (BD). PCDs, RLDs and BD are population-based data sources, while HD contains non-representative samples of the respective geographic catchment area. Average look-back time available in each data source at the beginning of the EMIF Project varied from 3 to 15 years.

Identification of Type 2 diabetes

A list of standard algorithms (***component algorithms***) for the identification of T2DM from the selected data sources was created. Each component algorithm was based on records **from one specific data domain among**: **diagnoses** (DIAG), **drug prescription** (DRUG), **utilization of a diagnostic test** (TEST) or **laboratory results** (LABVAL). For this purpose, two sources of knowledge were leveraged and integrated: a central expert-based clinical and operational definition of T2DM (top-down approach) and existing local expertise (bottom-up approach) provided by database experts. The Unified Medical Language System (UMLS) was used for **semantic harmonization** of coding systems: pertinent medical concepts were identified and projected to local terminologies (e.g. ICD9CM, ICD10, READ, ICPC for DIAG; ATC for DRUG; national coding systems for TEST and unit of measurement LABVAL). The resulting list of local codes and string was refined through an iterative process involving local experts' feedback.

Different logical combinations of components were tested by local experts which chose the preferred strategy for their data source (***recommended composite algorithm***) and provided a **comment as reusable knowledge**. Considering subjects 16+, all the person-time available at the index date (1st Jan 2012 for PCDs, RLDs and HD, 1st Jan 2009 for BD) was used in the case-identification algorithms.

Results

The EMIF-Platform provided aggregated health data on around 12 million European citizens. The component algorithms used for T2DM identification from the selected data sources are reported in Table 1. An example of the terminology mapping outcome is shown in Table 2. All recommended composite algorithms used at least one DIAG-based component as inclusion criteria, except for RLD-N that adopted a strategy based on DRUG only (Figure 1). DIAG-based inclusion criteria contributed for 93-100% of the total case population in PCD (Table 2), 100% in both BD and HD and 15-73% in RLD (Table 3). In RLDs, DRUG-based components identified from 81% to 100% of the respective total case population, and from 58% to 83% in PCDs. The population of cases identified through hypoglycaemic drug use (T2DM_ORAL) was the most homogeneous population across the 6 population-based data sources (PCDs and RLDs) (Figure 2).

Conclusions

Our standardization approach allowed to benchmark the results obtained form each extracted component algorithms across heterogeneous and otherwise non-comparable data sources. It also **provided insight into the total population of patients identified as cases in each specific data source** using the relevant composite strategy recommended by local experts. For instance, our results showed that in PCDs, BD and HD cases of T2DM were mainly identified through diagnoses: however, while in PCDs and BD cases are probably representative of patients with T2DM in the corresponding source population, cases identified in HD are expected to be more severe (i.e hospitalized). In RLDs DRUG-based components contributed most to the total case population. Indeed, such components do miss T2DM cases who are not treated with drugs and they may possibly misclassify T2DM with other diseases for which the same drugs can be used. Based on these considerations, **investigators and local experts could consider to change their preferred identification strategy** according to the type of study question or sensitivity analysis: if specificity is important, they may switch to DIAG-based identification strategies, at the expenses of sensitivity; if sensitivity is important, they may add other inclusion criteria, like LABVAL or TEST; if homogeneity across different data sources is important, they may agree to adopt a DRUG-based strategy. Notably, this data derivation process records *a priori* knowledge from each participating data source thus, whenever a study involving T2DM is designed, sensitivity analysis can be planned to discuss possible heterogeneity of study results. Ultimately, this data derivation process can be applied to any other event of interest.

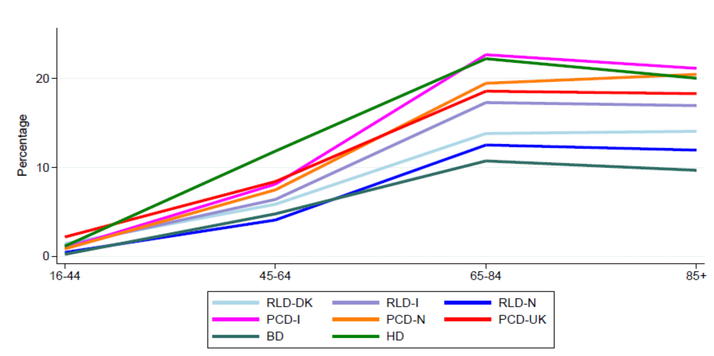
Table 1. List of standard component algorithms for the identification of type 2 diabetes cases from the selected data sources.

Algorithm acronym	Name	Description	Selection rules	Rules to identify subjects	Rules to identify date
T2DM_DIAG_PC	Diagnosis in primary care	Patients who have at least one diagnosis recorded in a primary care setting	(Diabetes type 2) occurs in [diagnosis fields] of [tables collected during primary care]	all subjects such that selection rule holds once or more	date of first record
T2DM_DIAG_SC	Diagnosis in secondary care	Patients who have at least one diagnosis recorded in a secondary care setting	(Diabetes type 2) occurs in [diagnosis fields] of [tables collected during secondary care]	all subjects such that selection rule holds once or more	date of first record
T2DM_DIAG_INP	Diagnosis in inpatient care	Patients who have at least one diagnosis recorded during a hospital admission	(Diabetes type 2) occurs in [diagnosis fields] of [tables collected during inpatient care]	all subjects such that selection rule holds once or more	date of first record
T2DM_DIAG_OTH	Diagnosis from other sources	Patients who have at least one diagnosis recorded in a setting other than primary, secondary or inpatient care	(Diabetes type 2) occurs in [diagnosis fields] of [tables collected in other settings]	all subjects such that selection rule holds once or more	date of first record
T2DM_DMUNSPEC	Diagnosis of unspecified DM in any healthcare setting	Patients who have at least one diagnosis of unspecified diabetes recorded in any care setting	(Diabetes unspecified) occurs in [diagnosis fields] of [any table collecting diagnoses]	all subjects such that selection rule holds once or more	date of first record
T2DM_DM1	Diagnosis of T1DM in any healthcare setting	Patients who have at least one diagnosis of type 1 diabetes recorded in any care setting	(Diabetes mellitus type I) occurs in [diagnosis fields] of [any table collecting diagnoses]	all subjects such that selection rule holds once or more	date of first record
T2DM_EXCL	Diagnosis of other conditions causing hyperglycemia, in any healthcare setting	Patients who have at least one diagnosis of conditions causing hyperglycemia other than T2DM and T1DM recorded in any care setting	((Metabolic problems around pregnancy) OR (Metabolic/pancreatic problems, non type 2 diabetes) OR (Polycystic Ovary Syndrome)) occurs in [diagnosis fields] of [any table collecting diagnoses]	all subjects such that selection rule holds once or more	date of first record
T2DM_ORAL	Oral antidiabetic utilization	Patients who have at least two prescriptions of oral antidiabetic in a calendar year	(Drugs used in diabetes, excl insulin) occurs in [ATC field] of [drugs tables]	all subjects such that selection rule holds twice or more in a year	date of second record
T2DM_ORAL_ONE	Oral antidiabetic utilization	Patients who have at least one prescription of oral antidiabetic in a calendar year	(Drugs used in diabetes, excl insulin) occurs in [ATC field] of [drugs tables]	all subjects such that selection rule holds once or more	date of first record
T2DM_INSULIN	Insulin utilization	Patients who have at least two prescriptions of insulin in a calendar year	(Insulins and analogues) occurs in [ATC field] of [drugs tables]	all subjects such that selection rule holds twice or more in a year	date of second record
T2DM_INSULIN_ONE	Insulin utilization	Patients who have at least one prescription of insulin in a calendar year	(Insulins and analogues) occurs in [ATC field] of [drugs tables]	all subjects such that selection rule holds once or more	date of first record
T2DM_LABVAL_TWO	Two glycated hemoglobin values higher than threshold	Patients who have at least two results recorded from a glycated hemoglobin test which is higher than 6.5% (48 mmol/mol)	(Glycated Haemoglobin) occurs in [code of test field] of [tables collecting laboratory test results] AND [result field] of the same record is higher than 6.5% (or 48 mmol/mol, according to unit of measurement adopted in the table)	all subjects such that selection rule holds twice or more	date of second record
T2DM_FAST_GLUC_TWO	Two fasting plasma glucose measurements higher than threshold	Patients who have at least two results recorded from a fasting plasma glucose measurement which is higher than 126 mg/dl)	(Fast Gluc) occurs in [code of test field] of [tables collecting laboratory test results] AND [result field] of the same record is higher than 126 mg/dl	all subjects such that selection rule holds twice or more	date of second record
T2DM_LCURVE_GLUC_TWO	Two glucose tolerance tests higher than threshold	Patients who have at least two results recorded from a glucose tolerance test higher than 200 mg/dl)	(LcurveGluc) occurs in [code of test field] of [tables collecting laboratory test results] AND [result field] of the same record is higher than 200 mg/dl	all subjects such that selection rule holds twice or more	date of second record
T2DM_TEST_GLUCO5_1YR	Repeated measurement of glucose in one year	Patients who have a fifth blood glucose measurement within 1 year	(Blood glucose measurement) occurs in [code of test field] of [tables collecting laboratory test results or dispensings]	at least 5 in a year	date of the fifth record within a year
T2DM_GLUCO2_PYEAR_5YR	Repeated measurement of glucose in one year	Patients who have two blood glucose measurements per year for 5 consecutive years	(Blood glucose measurement) occurs in [code of test field] of [tables collecting laboratory test results or dispensings]	at least 2 per year in 5 consecutive years	date of the second record in the 5th year

Table 2. Example of terminology mapping output: diagnosis codes and free text corresponding to the concept "type 2 diabetes"

Unified Medical Language System - Concept Unique identifiers		Local terminologies				
		ICD9CM	ICD10	ICPC2P	RCD	Free text in PCD-N
C0011860	Diabetes Mellitus, Non-Insulin-Dependent		E11	T90005, T90007, T90009	X40J5	('T90') AND ('type 2')
C0375113	type II diabetes mellitus [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled, without mention of complication	250.00				
C0375115	type II diabetes mellitus [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled, without mention of complication	250.02				
C0375117	type II diabetes mellitus [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled, with ketoacidosis	250.10				
C0375119	type II diabetes mellitus [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled, with ketoacidosis	250.12				
C0375122	Diabetes mellitus, type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type with hyperosmolarity, not stated as uncontrolled	250.20				
C0375124	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type with hyperosmolarity, uncontrolled	250.22				
C0375126	Diabetes mellitus, type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled	250.30				
C0375128	Diabetes mellitus with other coma, type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled	250.32				
C0375130	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled, with renal manifestations	250.40				
C0375132	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled, with renal manifestations	250.42				
C0375134	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled, with ophthalmic manifestations	250.50				
C0376128	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled, with ophthalmic manifestations	250.52				
C0375137	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled, with neurological manifestations	250.60				
C0375139	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled, with neurological manifestations	250.62				
C0375141	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled, with peripheral circulatory disorders	250.70				
C0375143	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled, with peripheral circulatory disorders	250.72				
C0375145	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled, with other specified manifestations	250.80				
C0375147	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled, with other specified manifestations	250.82				
C0375149	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, not stated as uncontrolled, with unspecified complication	250.90				
C0375151	Diabetes mellitus type II [non-insulin dependent type] [NIDDM type] [adult-onset type] or unspecified type, uncontrolled, with unspecified complication	250.92				

Figure 1. Recommended composite algorithms: percentage of the total data base population identified per age band



Data Source	Recommended composite algorithm	Comment of the local expert	Sensitivity	PPV
PCD-I	(T2DM_DMUNSPEC **OR** T2DM_LABVAL_TWO **OR** T2DM_FAST_GLUC_TWO **OR** T2DM_LCURVE_GLUC_TWO) **AND NOT** (T2DM_DM1))	The chosen composite algorithm was validated in HSD in a study that is undergoing publication and found very high PPV (around 100% in the validation sample). Due to the nature of the data source and the very broad algorithm, sensitivity must be very high as well.	≥ .9	≥ .9
PCD-N	T2DM_DIAG_PC **OR** T2DM_ORAL_ONE	Diagnosis of T2DM is not always recorded, so we included subjects who utilize non-insulin antidiabetic drugs (ORAL). Beware that some subjects have a record of DM1, because some GPs record this diagnosis to indicate insulin dependence also in subjects with T2DM.	≥ .9	≥ .9
PCD-UK	T2DM_DIAG_PC	adding any other strategy (LABVAL, ORAL, INSULIN...) does not add almost anybody	≥ .9	≥ .9
RLD-DK	(T2DM_DIAG_INP **OR** T2DM_DIAG_SC) **OR** ((T2DM_TEST_GLUCO5_1YR **OR** T2DM_ORAL **OR** T2DM_INSULIN **OR** T2DM_GLUCO2_PYEAR_5YRS)**AND NOT** (T2DM_DM1 **OR** T2DM_EXCL))	The chosen strategy to identify T2DM cases is similar to the strategy which is regularly used in AUH to identify cases of unspecified diabetes. This is the strategy of the Danish National Diabetes registry, which has been repeatedly validated. A recent study published in 2015 estimated that sensitivity and PPV are 95 and 80%, respectively. When adapting this algorithm to the case of T2DM we decided to change some elements of the validated strategy. The main differences are: we used type 1 diabetes diagnoses as exclusion criteria, we used diagnoses of T2DM rather than diagnoses of unspecified diabetes, we avoided chiropody for diabetics as inclusion criteria and we did not exclude cases of gestational diabetes. As for insulin and other antidiabetic drugs, we used two prescription in one year as inclusion criteria rather than two prescription recorded at any time. The expected sensitivity and PPV of the chosen algorithm are possibly slightly lower but still very close to those of the validated algorithm.	> 7 and < .9	> 7 and < .9
RLD-I	T2DM_DIAG_INP **OR** T2DM_ORAL **OR** T2DM_INSULIN **OR** T2DM_DIAG_OTH	From a validation study, the sensitivity of this algorithm is 76% and PPV 86%, excluding subjects with a record of T1 diabetes does not improve the extraction validity. T2DM_DIAG_OTH refers to an Italian table: disease-specific exemptions from copayment to healthcare	> 7 and < .9	> 7 and < .9
RLD-N	T2DM_ORAL	The chosen strategy to identify type 2 diabetes patients (T2DM_ORAL) has a very high PPV (95%). The reason we have decided to use only this method and not include other components we extracted is that the identification of T2DM patients based on the use of oral antidiabetics is extensively tested and validated within GPs. We are aware that this approach is limited to identify treated patients only, but this is deliberate. Although a patient could be classified as a T2DM patient given the strict clinical definition that was chosen in this process (i.e. HbA1c above a certain threshold), the fact that the patient is not treated begs the question whether it should be classified as a prevalent patient. The sensitivity of this strategy with respect to the wider class of patients matching the clinical definition is between 70 and 90%.	> 7 and < .9	≥ .9
BD	T2DM_DIAG_PC	Using any other extracted component algorithm as additional inclusion criteria (i.e. LABVAL, INSULIN_ONE) do not add any patients if those with a concomitant diagnosis of type 1 diabetes are excluded.	≥ .9	≥ .9
HD	T2DM_DIAG_INP **AND NOT** T2DM_DM1	The chosen algorithm uses as inclusion criterion inpatients diagnoses of T2DM and exclude patients with diagnosis of type 1 diabetes. Other components such as T2DM_EXCL and T2DM_DIABPREG were not considered because did not significantly affect the population of cases identified (total patients excluded=46 of 2,483 - that is 1,9%).	≥ .9	≥ .9

Figure 2. Percentage of data base population identified by individual component algorithms per age band: four examples

Table 3. Impact of each component algorithms on the population of cases retrieved through recommended composite algorithms

COMPONENT ALGORITHMS (B)°		RECOMMENDED COMPOSITE ALGORITHMS (A)							
		RLD-I	RLD-DK	RLD-N	PCD-UK	PCD-N	PCD-I	BD	HD
	N	3391177	1372883	1405220	3278013	992924	945691	22430	15713
	N in A	254045	77616	57712	253197	67096	81658	779	2466
	% of A in N	7.5	5.7	4.1	7.7	6.8	8.6	3.5	15.7
T2DM_DIAG_PC	N in B	0	0	0	253197	62191	43438	779	0
≥1 diagnosis	% of B in A	-	-	-	100.0%	92.7%	52.6%	100.0%	-
from primary care	PR if B added	-	-	-	+0.0%	+0.0%	+0.6%	+0.0%	-
T2DM_DIAG_INP	N in B	95303	27887	13098	0	0	0	0	2520
≥1 T2DM diagnosis	% of B in A	37.5%	35.9%	15.1%	-	-	-	-	100.0%
from inpatient care	PR if B added	+0.0%	+0.0%	+7.6%	-	-	-	-	+2.2%
T2DM_DIAG_SC	N in B	0	35744	0	0	0	0	0	0
≥1 T2DM diagnosis	% of B in A	-	46.1%	-	-	-	-	-	-
from secondary care	PR if B added	-	+0.0%	-	-	-	-	-	-
T2DM_DMUNSPEC	N in B	191999	0	0	0	0	79035	0	0
≥1 unspecified diabetes diagnosis	% of B in A	73.2%	-	-	-	-	94.3%	-	-
from any healthcare setting	PR if B added	+2.4%	-	-	-	-	+2.5%	-	-
DM_DIAG_OTH	N in B	149806	0	0	0	0	0	0	0
≥1 unspecified diabetes diagnosis	% of B in A	59.0%	-	-	-	-	-	-	-
from co-payment exemption	PR if B added	+0.0%	-	-	-	-	-	-	-
T2DM_DM1	N in B	18147	17896	0	0	8816	2050	164	78
≥1 type 1 diabetes diagnoses	% of B in A	6.9%	18.1%	-	-	8.8%	0.0%	2.8%	0.0%
from any healthcare setting	PR if B added	+0.2%	+4.9%	-	-	+4.3%	+2.5%	+18.2%	+3.2%
T2DM_EXCL	N in B	13741	7895	2904	0	0	5782	0	78
≥1 diagnoses of other types of	% of B in A	1.1%	1.8%	1.5%	-	-	0.3%	-	1.7%
diabetes or glucose intolerance	PR if B added	+4.3%	+8.3%	+3.5%	-	-	+6.8%	-	+1.5%
T2DM DIAG PC OR **T2DM DIAG INP** OR **T2DM DIAG SC** OR **T2DM DMUNSPEC** OR **T2DM DIAG OTH**	N in B	191999	43622	13098	253197	62191	79035	779	2520
	% of B in A	73.2%	56.2%	15.1%	100.0%	92.7%	94.3%	100.0%	100.0%
	PR if B added	+2.4%	+0.0%	+7.6%	+0.0%	+0.0%	+2.5%	+0.0%	+2.2%
T2DM_INSULIN	N in B	45522	22074	21192	41019	15020	11607	-	0
≥2 prescriptions of	% of B in A	17.9%	25.4%	25.8%	16.1%	19.0%	12.3%	-	-
insulin in one year	PR if B added	+0.0%	+3.0%	+10.9%	+0.1%	+3.4%	+2.0%	-	-
T2DM_INSULIN_ONE	N in B	62341	23319	0	0	17719	0	18	0
≥1 prescription of	% of B in A	21.2%	26.5%	-	-	22.0%	-	1.5%	-
insulin in one year	PR if B added	+3.4%	+3.6%	-	-	+4.4%	-	+0.8%	-
T2DM_ORAL	N in B	216338	57153	57712	136370	51589	45624	-	0
≥2 prescriptions of	% of B in A	85.2%	71.0%	100.0%	51.7%	76.9%	53.0%	-	-
NIAD in one year	PR if B added	+0.0%	+2.7%	+0.0%	+2.1%	+0.0%	+2.9%	-	-
T2DM_ORAL_ONE	N in B	273952	61604	0	0	54181	62110	45	0
≥1 prescription of NIAD	% of B in A	87.5%	72.7%	-	-	80.8%	70.6%	5.8%	-
in one year	PR if B added	+20.3%	+6.7%	-	-	+0.0%	+5.4%	+0.0%	-
T2DM_INSULIN OR **T2DM_INSULIN_ONE** OR **T2DM_ORAL** OR **T2DM ORAL_ONE**	N in B	295676	70405	64016	151576	58355	65076	40	0
	% of B in A	93.0%	81.1%	100.0%	57.7%	82.6%	73.1%	50.0%	-
	PR if B added	+23.4%	+9.6%	+10.9%	+2.2%	+4.4%	+6.6%	+4.1%	-
T2DM_FAST_GLUC_TWO	N in B	0	0	0	0	0	32153	0	0
≥2 fasting glucose values	% of B in A	-	-	-	-	-	38.6%	-	-
>126mg/dl	PR if B added	-	-	-	-	-	+0.8%	-	-
T2DM_LABVAL_TWO	N in B	0	0	62400	0	44271	20196	0	0
≥2 glycated	% of B in A	-	-	65.1%	-	63.6%	24.1%	-	-
hemoglobin value	PR if B added	-	-	+43.0%	-	+2.4%	+0.7%	-	-
T2DM_LCURVE_GLUC_TWO	N in B	0	0	0	0	0	32	0	0
≥2 glucose tolerance test	% of B in A	-	-	-	-	-	0.0%	-	-
values >200mg/dl	PR if B added	-	-	-	-	-	+0.0%	-	-
T2DM_FAST_GLUC_TWO OR **T2DM_LABVAL_TWO** OR **T2DM_LCURVE_GLUC_TWO**	N in B	0	0	62400	0	44271	38764	0	0
	% of B in A	-	-	65.1%	-	63.6%	46.5	-	-
	PR if B added	-	-	+43.0%	-	+2.4%	+1.0%	-	-
T2DM_TEST_GLUCO5_1YR	N in B	266940	16999	0	0	0	0	0	0
≥5 glycated hemoglobin tests	% of B in A	45.8%	21.6%	-	-	-	-	-	-
in 1 year	PR if B added	+59.3%	+0.3%	-	-	-	-	-	-
T2DM_GLUCO2_PYEAR_5YRS	N in B	172784	28583	0	0	0	0	0	0
≥2 glycated	% of B in A	32.6%	36.1%	-	-	-	-	-	-
	PR if B added	+35.4%	+0.7%	-	-	-	-	-	-
T2DM_TEST_GLUCO5_1YR OR **T2DM_GLUCO2_PYEAR_5YRS**	N in B	335466	34801	0	0	0	0	0	0
	% of B in A	52.8%	44.1%	-	-	-	-	-	-
	PR if B added	79.2%	+0.8%	-	-	-	-	-	-

Since patients can be identified by more than one component algorithms, percentages may overlap.
°Only standardized component algorithms that were included in at least one final composite algorithm are shown in the table.
A= recommended composite algorithm
B= component algorithm(s)
N= number of active patients at aged 16+ at index date
PR= percentage ratio of patients in A or B in N with respect to the percentage of A in N
Grey cells correspond to component algorithms that were included in the relevant recommended composite algorithm.

Funding acknowledgement:
This work has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking (EMIF grant n° 115372)

31st International Conference on Pharmacoepidemiology & Therapeutic Risk Management
22-26 August 2015 – Boston, MA USA