

Contents lists available at ScienceDirect

Vaccine





Age-specific vaccination coverage estimates for influenza, human papillomavirus and measles containing vaccines from seven population-based healthcare databases from four EU countries – The ADVANCE project



Toon Braeye ^{a,b,*}, Hanne-Dorthe Emborg ^c, Ana Llorente-García ^d, Consuelo Huerta ^d, Elisa Martín-Merino ^d, Talita Duarte-Salles ^e, Giorgia Danieli ^{f,g}, Lara Tramontan ^{f,g}, Daniel Weibel ^h, Chris McGee ⁱ, Marco Villa ^j, Rosa Gini ^k, Matti Lehtinen ^l, Lina Titievsky ^m, Miriam Sturkenboom ^{n,o,p}

- ^a Sciensano, Brussels, Belgium
- ^b University of Hasselt, Hasselt, Belgium
- ^c Statens Serum Institut, Copenhagen, Denmark
- ^d BIFAP Database, Spanish Agency of Medicines and Medical Devices, Madrid, Spain
- e Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain
- f Consorzio Arsenàl.IT, Veneto Region, Italy
- Epidemiological Information for Clinical Research from an Italian Network of Family Paediatricians (PEDIANET), Padova, Italy
- ^h Erasmus University Medical Center, Rotterdam, The Netherlands
- ⁱ University of Surrey, Surrey, UK
- ^j ATS della Val Padana, Cremona, Italy
- ^k Agenzia Regionale di Sanita Della Toscana, Florence, Italy
- ¹University of Tampere, Tampere, Finland
- ^m Epidemiology, Pfizer, NY, United States
- ⁿ P95, Heverlee, Belgium
- ° VACCINE.GRID, Basel, Switzerland
- ^p University Medical Center Utrecht, Julius Global Health, Utrecht, The Netherlands

$A\ R\ T\ I\ C\ L\ E\quad I\ N\ F\ O$

Article history:
Received 16 July 2019
Received in revised form 20 February 2020
Accepted 28 February 2020
Available online 12 March 2020

Keywords: Vaccination coverage Influenza vaccines Papillomavirus vaccines Measles vaccines Probability

ABSTRACT

Background: The Accelerated Development of VAccine beNefit-risk Collaboration in Europe (ADVANCE) is a public-private collaboration aiming to develop and test a system for rapid benefit-risk monitoring of vaccines using existing healthcare databases in Europe. We estimated vaccine coverage from electronic healthcare databases as part of a fit-for-purpose assessment for vaccine benefit-risk studies.

Methods: A retrospective dynamic cohort study was conducted through a distributed network approach. Coverage with measles-vaccine for birth year 2006, human papillomavirus (HPV)-vaccine for birth years 1990–2000 and influenza-vaccine for birth years 1920–1950 was estimated using period-prevalence and inverse probability weighting methods. Seven databases from four countries participated: Italy (Pedianet, Val Padana), Spain (BIFAP, SIDIAP), UK (RCGP-RSC, THIN), Denmark (SSI/AUH). Database access providers extracted the data, transformed it into a common structure and ran an R-script locally. The created output tables were shared and pooled at a central server.

Results: The total study population comprised 274,616 persons for measles-vaccine, 2,011,666 persons for HPV-vaccine and 14,904,033 persons for influenza-vaccine. Measles-vaccine coverage varied from 84.3% (Denmark) to 96.5% (Italy, Val Padana) for the first dose and from 82.8% (Italy, Val Padana) to 90.9% (UK) for the second dose at the age of 7 years. The HPV-vaccine coverage, aggregated over birth years 1997–2000, ranged from 60% (UK) to 88.3% (Denmark) at the age of 15 years. The influenza-vaccine coverage for the influenza seasons from 2009 to 2015 for persons aged 65 years and more was roughly stable around 43% in Denmark and around 68% in the UK while a decrease from 58 to 50% was observed in Catalonia (Spain).

^{*} Corresponding author at: Sciensano, Rue Juliette Wytsman 14, 1050 Ixelles, Belgium. E-mail address: toon.braeye@sciensano.be (T. Braeye).

Conclusions: We obtained detailed, age-specific coverage estimates though a common procedure. We discussed between database comparability and comparability to published national estimates.

© 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The impact of a vaccine on the burden of a disease in populations differs from an assessment in individuals. Because of effects like herd immunity, the population coverage should be included in benefit-risk vaccine studies [1]. While several sources such as surveys, social security data or health insurance claims data [2] can be used to obtain coverage estimates, electronic healthcare records (eHR) are likely to gain importance. Their growing importance as source to rapidly provide evidence on aspects of vaccine coverage, benefits and risks has multiple reasons: I) eHR are less likely associated with biases, such as recall or selective sampling bias, that are important for other sources [3], II) eHR potentially allow for risk group specific analysis, a necessary feature since vaccine recommendations can be risk group specific and III) eHR allow for age-specific, near real time coverage estimation with a high positive predictive value at relatively low cost [4–6]. Despite these advantages, real world applications have reported issues with incompleteness and misclassification [7-9]. These concerns and the fact that there are many different kinds of eHR-databases require a 'fit for purpose'-assessment prior to their use in benefit-risk hypothesis testing.

Most European countries produce regional or national vaccine coverage estimates, but databases and methods used to estimate vaccination coverage vary widely and comparability is challenging at an international level [10-12]. Several projects and organisations, such as the WHO Centralized Information System for Infectious Diseases (CISID) [13] and Vaccine European New Integrated Collaboration Effort (VENICE) [14] aim to improve the quality and comparability of vaccine coverage estimation by defining and implementing standards. These projects have been in place for several years and have established robust databases. An alternative approach to obtain comparable coverage estimates from multiple countries is the standardized collection of data using surveys [15,16]. The ADVANCE project differed from these existing approaches as we did not require our partners to present coverage estimates themselves nor did we collect data ourselves. Instead, participating databases were asked to run a script that transformed available eHRs into vaccination coverage estimates locally. This insured that the steps for data cleaning, transformation and analysis were identical between databases. In a previous paper, we investigated methods to deal with one source of incompleteness encountered in eHR-databases; unregistered vaccinations because of incomplete follow-up [17]. From that paper we selected two methods that take person-time into account when estimating coverage.

The Accelerated Development of VAccine beNefit-risk Collaboration in Europe (ADVANCE) project was a public private collaboration aiming to develop and test a system for rapid benefit-risk monitoring of vaccines using existing healthcare databases in Europe. For further reading on the ADVANCE project and its code of conduct we refer to http://www.advance-vaccines.eu/. Several studies were conducted to demonstrate the ability to generate evidence on pertussis vaccine coverage, benefits and risks [18]. The results presented here were part of a second round of feasibility assessments to create readiness to study additional vaccines beyond pertussis. In this paper we estimated the coverage with measles, HPV and influenza-vaccines in participating eHR-databases.

2. Material and methods

2.1. Design and setting

We conducted a retrospective dynamic cohort study using a distributed network approach.

2.2. Description of participating databases

Seven European healthcare databases participated in this study (Table 1). The general characteristics of the databases have been described in more detail before [19]. All databases comprised electronic health records. There was a distinction between regional (BIFAP, SIDIAP, PEDIANET, Val Padana) and national databases (THIN, RCGP-RSC, SSI/AUH). There was also a distinction between databases that collected a subset (BIFAP, SIDIAP, PEDIANET, THIN, RCGP-RSC) or aimed to collect all data (Val Padana, SSI/AUH) from a given area. In PEDIANET only children whose parents consented linkage to the Veneto regional vaccination registry were included. Finally, data was collected mainly from primary care healthcare services, except for SSI/AUH and Val Padana where linkage between registries was performed.

2.3. Vaccinations

The vaccines of interest in this assessment comprised measlescontaining vaccines, HPV-vaccines and seasonal influenza vaccines. Vaccination schedules for these vaccines differed between the countries included in this study: Denmark, Italy, United Kingdom and Spain (Table 2). Vaccination information was extracted from the eHR-databases locally and transformed into a common vaccination file which comprised the following variables: a patient identifier, a vaccine type (coded as the disease against which the vaccine protected), Anatomical Therapeutic Chemical (ATC-code, according to WHO), a brand name, a recorded dose number and/ or a derived dose number (derivation based on age and sequence) and a date of vaccination. The brand and ATC-code could be left empty. Val Padana and RCGP-RSC provided a derived dose (derived from the sequence of observed doses). The other databases provided a recorded dose. To determine the seasonal influenza vaccine coverage, we looked at the date of vaccination. For example, an influenza vaccination registered between 1 July 2010 and 30 June 2011 was included in the 2010/11 seasonal coverage estimation.

2.4. Source and study population

The source population for this assessment comprised all subjects registered in the databases listed above. Each database provided a start- and end-date of follow-up for each person. Reentry into the database, providing several start- or end-dates per person, was not permitted. Other variables available at the individual level in the local data were gender and birth date. Prior population characterization has shown that the population captured by the databases reflected the age and gender distribution in the countries [19].

For each vaccine, we defined study cohorts by birth years. The age at the start of follow-up was an additional inclusion criterion for measles and HPV-vaccine coverage estimation. This inclusion

Table 1Characteristics of the databases included in the fit-for-purpose.

Country	Denmark	Spain		Italy		United Kingdo	om
Name Type of organisation providing access	SSI/AUH Different public data holders	BIFAP Spanish regulatory agency	SIDIAP Public research organisation	PEDIANET Private organisation; vaccines from public health	Val Padana Local public health agency	THIN Academic License holder (Erasmus MC)	RCGP-RSC Charity
Origin of data	Hospital discharge diagnoses linked to population and vaccination registries. National health care	Family paediatricians and general practitioners medical records	Family paediatricians and general practitioners medical records	Family paediatricians medical records linked to regional vaccination register	Hospitalisation discharge diagnoses linked to population and vaccination registries	General practitioners medical records	General practitioners medical records
Geographic spread	National	Multiregional:9 out of 17 regions	Catalonia Region	Sample from Veneto Region	Regional, province	National sample	National sample

Table 2
Overview of the recommended age of vaccination by country and vaccine, data obtained from eCDC [20].

Country	Databases	1 ^e Measles	2 ^e Measles	1 ^e HPV	Influenza recommended to persons > 65 years
Denmark	SSI/AUH	15 m	4y	12y	Yes
Italy	PEDIANET - Val Padana	13-15 m	5-6y	12y	Yes
Spain	BIFAP – SIDIAP	12 m	3-4y	11-14y	Yes
UK	RCGP-RSC - THIN	12 m	3y	12-13y	Yes

criterion determined the amount of left censoring. Left censoring was defined as unregistered vaccinations because follow-up had not started at the time of vaccination. We estimated age-specific measles-vaccine coverage for children born in 2006. Only children who started follow-up before the age of 100 weeks were included in the measles analysis. For the estimation of coverage with HPV-vaccine, the study population comprised all female children born between 1990 and 2000 who started follow-up before the age of nine years. To estimate influenza-vaccine coverage in the elderly (65 years and older) for the influenza seasons between 2009 and 2015, the study population comprised all persons born between 1920 and 1950.

2.5. Distributed data processing

In the distributed approach, data access providers extracted and transformed the data into the common data model (CDM) comprising the individual-level vaccination and population files described above. An R-script was sent to the data access providers. The Rscript conducted cleaning, application of inclusion and exclusion criteria, data collection for the attrition tables, transformation and coverage estimation. Cleaning of the study population consisted of removing incomplete records, incorrect records (e.g. a negative period of follow-up) and duplicates. Duplicated vaccination records were defined as having the 'same dose, same vaccine component' and 'same date, same vaccine component'. After linking the population file to the vaccination file, vaccinations occurring outside of an individual's start- and end-date of follow-up were censored. The attrition tables provided information on: the number of persons by birth year and their time in follow-up, the number of persons excluded because their age at the start of follow-up exceeded the predefined limit, the number of vaccination records.

The tables that the R-script generated contained age-specific coverage estimates by method, week of age, birth year, vaccine component and dose. Bootstrapped 95% confidence intervals were obtained for the measles-vaccine by sampling from the input-records randomly with replacement. A total of 1000 samples were

taken for the calculation of each bootstrapped confidence interval. Tables with aggregated data were shared to a remote research environment through a secure file transfer protocol.

Central to our approach is that a single R-script was provided to all databases. The script extracted and analysed the data available for the previously described study populations. Not all databases were able to provide all estimates. Pedianet could not provide information on HPV-vaccine, since only birth years 2006 and 2007 were linked to vaccination registries, nor on influenza in adults, as this is a paediatric database. Val Padana did not contribute to the influenza-vaccine coverage estimation because the birth years under investigation were not present in the database. Several databases could not provide HPV-vaccine coverage estimates for the birth years 1990–1993. THIN did not provide estimates for the second dose measles vaccine. THIN and BIFAP did not provide all seasonal influenza-vaccines estimates.

2.6. Coverage estimation

Coverage was estimated by birth year over age in weeks. For example for persons born in 2006, we estimated the coverage with the first dose measles-vaccine at 1 week of age, 2 weeks of age, etc. Age in weeks was calculated as $\frac{birthdate-studydate}{7}$ rounded down. The number of persons in follow-up for at least one day during an age in weeks was counted. We further counted the number of persons who received a vaccination during that week, and those who had a registered vaccination prior to that age in weeks. From these counts we calculated two coverage estimates: a period prevalence $(PP.fu_i)$ estimate and an inverse probability weighted (IPW) estimate.

The $PP.fu_i$ for age in weeks i is estimated as:

$$PP.fu_i = \frac{N \, vaccinated \& in FU_i}{Nin FU_i}$$

With $PP.fu_i$: the coverage estimated by the PP.fu-method during week i, $Nvaccinated\&inFU_i$ the number of persons in follow-up during week i who have been vaccinated prior to or during week and $NinFU_i$: the total number of persons in follow-up during week i.

$$FU_{proportion,i} = \frac{NinFU_i}{NinFU}$$

With $FU_{proportion,i}$: the proportion of persons in follow-up during week i, $NinFU_i$: the total number of persons in follow-up during week i and NinFU: the total number of persons in the birth cohort.

$$Vacc_{IPW,i} = \frac{Vacc_{observed,i}}{FU_{proportion,i}}$$

With $Vacc_{IPW,i}$: the estimated number of vaccination during week i, $Vacc_{observed,i}$: the observed number of vaccinations during week i and $FU_{proportion,i}$: the proportion of persons in follow-up during week i.

$$IPW_i = \frac{\sum_{0 \to i} Vacc_{IPW,i}}{NinFU}$$

With IPW_i : the coverage estimated by the IPW-method during week i, $Vacc_{IPW,i}$: the estimated number of vaccination during week i and NinFU: the total number of persons in the birth cohort.

A simple example to illustrate the method; when 50% of the study cohort was in follow-up during age week 3 and 100 vaccinations were registered during this week, we assumed that 200 vaccinations were administered at 3 weeks of age (100 registered + 100 censored).

The PP.fu-method relies on the assumption that the age-specific coverage estimated from the part of the population in follow-up at any age in weeks represent the age-specific coverage of the population. The IPW-method relies on the assumption that the proportion of persons in follow-up receiving a vaccine during a certain age in weeks equals the proportion of persons not in follow-up receiving a vaccine. Since this assumption is likely violated for the influenza-vaccine study population, as in older age groups death is a common cause of loss to follow-up, the IPW-method was not applied to influenza-vaccine. A general summary of these assumptions is that with the PP.fu-method we assumed that the observed coverage equalled the study population coverage, while with the IPW-method we estimated the coverage. Both methods will deliver biased estimates when the probability of vaccination differs in and out of follow-up.

3. Results

3.1. Measles-vaccine coverage

For measles-vaccine, the seven databases contributed a total of 362,063 persons born in 2006, of which 274,616 were eligible for analysis. The exclusion of persons was due to entering the database at an age older than 100 weeks (N = 87,447). A total of 401,094 vaccine doses (dose 1 and dose 2) were recorded for this study population. The numbers by database can be found in the attrition tables in the supplementary file.

Fig. 1 shows the proportion of persons in the study population in follow-up from birth to seven years of age. The study population in Pedianet that had consented to be linked to the vaccine registry had complete follow-up. The proportion in follow-up in the national SSI/AUH-databases remained high. In the regional Val Padana and primary care databases the proportion of persons in follow-up was more dynamic; the proportion of persons in follow-up increased until the age-limit (100 weeks of age) set as additional inclusion criterion. After which, the proportion of persons in follow-up decreased. The decrease was substantially for the BIFAP, RCGP-RSC and THIN-databases.

Fig. 2 shows the age-specific coverage estimates for the first and second dose of measles-vaccines, using the PP.fu-method and the IPW-method. Table 3 shows the estimates at seven years of age. The start of uptake for the first dose of measles-vaccine ranged from 0.6 years (SIDIAP) to 1.2 years of age (SSI/AUH). After the

initial steep increase in coverage at one year of age, THIN, RCGP-RSC and SSI/AUH reported a continuing increase over time, representing vaccinations after the recommended age for vaccination. A second increase was seen at the age of five years in the Val Padanadatabase. At seven years, the IPW-estimate for the first dose ranged from 96.5% (Val Padana) to 84.3% (SSI/AUH).

For the second dose of the measles-vaccine only RCGP-RSC reached a coverage of 90% (IPW-estimate). The increase in coverage over age was less steep, stepwise for some databases, and the attained coverage for the second dose was below the coverage attained for the first dose. The minimum difference between the first and second dose coverage was 1.5% (BIFAP), the maximum difference was 13.7% (Val Padana). There were differences between countries with respect to the age at which the coverage with the second dose started increasing. For SIDIAP, a second dose coverage of around 30% was reached at the age of two years. At seven years the IPW-estimate for the second dose ranged from 90.9% (RCGP-RSC) to 80.0% (SSI/AUH).

3.2. HPV-vaccine coverage

The HPV study population of females born between 1990 and 2000 with follow-up before the age of nine years comprised 2,011,666 persons. The total number of first dose recorded HPV-vaccine registrations for the study population was 838,823. Several of the databases did not contribute to the birth cohorts from 1990 to 1993. For the 1990 birth cohorts to be included, the databases needed to have started follow-up by 1999 (because the 'age at start of follow-up'-inclusion criterion was set at nine years) and this information needed to be present in the input-files. An overview of the study population by database can be found in the attrition tables in the supplementary file.

The age at the start of HPV-vaccination differed between databases as did the coverage attained at the age of 15 years. E.g. in the UK (THIN & RCGP-RSC) HPV-vaccination started one birth year later than in Denmark (SSI/AUH) and Spain (BIFAP). SSI/AUH attained a HPV-vaccine coverage of >70% for birth year 1994 after which the coverage continued to increase. The other databases, except for RCGP-RSC and THIN, also reported an increasing coverage over the birth years included in this study. The attained HPV-vaccine first dose coverage at the age of 15 years for females born between 1997 and 2000 estimated with the IPW-method ranged between 60% (RCGP-RSC/THIN) and 88.3% (SSI/AUH) (Table 4) (Fig. 3). A figure with PP.fu-estimates is provided in the supplementary file.

3.3. Influenza-vaccine coverage

Five databases captured information on influenza-vaccines in the elderly. The number of persons included in the estimation increased from 3,686,640 persons for the influenza season 2009–2010 to 4,932,582 for the influenza season 2014–2015. A figure of the proportion of persons in follow-up per database per season and the attrition tables are provided in the supplementary file.

There were large between databases differences in the attained coverage. SSI/AUH reported coverages between 40 and 45% for all seasons, while THIN and RCGP-RSC reported coverages around 70%. There was a downward trend in coverage in recent seasons in the SIDIAP-database (from 58% in 2009/10 to 50% coverage in 2013/14 seasons). The other databases showed no clear timetrend in estimated coverage from the 2009–2010 season to the 2014–2015 season (Fig. 4).

Influenza-vaccine coverage increased with age for persons aged 65 years and more. The increase was followed by a decline in the oldest age groups. In the UK (RCGP-RSC and THIN) the highest coverage was reached in persons aged 77–80 years. In Spain (BIFAP

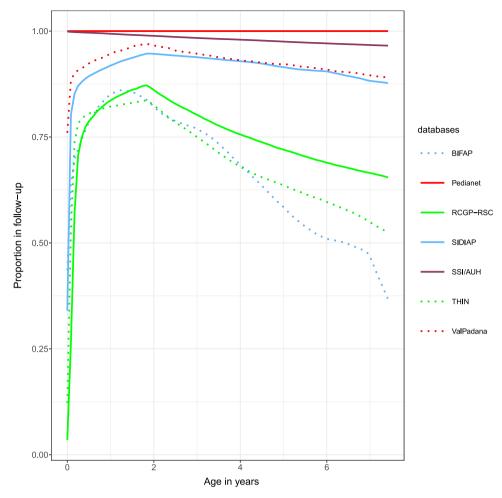


Fig. 1. Proportion of persons in follow-up, with start of follow-up before the age of 100 weeks and birth year 2006 over age in years, by database.

and SIDIAP) and Denmark (SSI/AUH) the highest coverage was reached at a later age, around 85–87 years. This pattern was seen over all studied influenza seasons. Due to the low number of persons in follow-up in the oldest age groups, the PP.fu-estimate became unstable (Fig. 5).

4. Discussion

This study aimed to assess part of the suitability of eHR-databases to participate in distributed vaccine studies. We also aimed to test the broad use of a common data model and a common analysis script. We found that, given that the data were available, age-specific, but also calendar year or season-specific coverage estimates could be obtained through this common procedure. Limitations however exist, both with respect to methods for coverage estimation and the estimates obtained.

4.1. Methods for coverage estimation

Previous research has pointed to the importance of both the total uptake and the timeliness [21]. We therefore selected two methods that allowed for age-specific estimates and could represent vaccinations before and after the recommended age at vaccination. Since left and right censoring because of incomplete follow-up were present in different proportions in several of the databases, the methods also had to be able to account for censored events. Censoring was especially present in databases linked to

general practice. The PP.fu- and IPW-method accounted for censoring in different ways. PP.fu and IPW-estimates were comparable when incompleteness in follow-up was limited. As there was no loss to follow-up in the Pedianet-database, the IPW and PP.fu-estimates were equal. The absence of incomplete follow-up in the Pedianet database is an artefact of the need to ask for consent to be linked to the immunization registry. Consent was only obtained for persons still registered in 2015 and born in 2006 and 2007.

A limitation of the PP.fu-method was the inability to account for left censoring; entering the database after vaccination caused underestimation [22,23]. As we had set the inclusion criterion for the first dose of measles-vaccine at a starting age of follow-up of 100 weeks (an age at which most children had already received a first dose of measles-vaccine) and because follow-up was dynamic in some databases, left censoring was likely. As a result, the PP.fuestimates for the first dose of measles-vaccine were lower than the IPW-estimates for some databases. As incompleteness increased, the difference between IPW and PP.fu-estimates became larger. It was largest for the THIN and BIFAP-database. If the 'age at the start of follow-up'-inclusion criterion was set before the recommended age of vaccination, it could avoid most left-censoring and therefore underestimation by the PP.fu-method. This happened for both the second dose of measles-vaccine and for the first dose of HPVvaccine. Overall, an 'age at the start of follow-up'-inclusion criterion can reduce left censoring, but it will also reduce the study population to a smaller sample and that sample might no longer be representative for the total population. This criterion and the

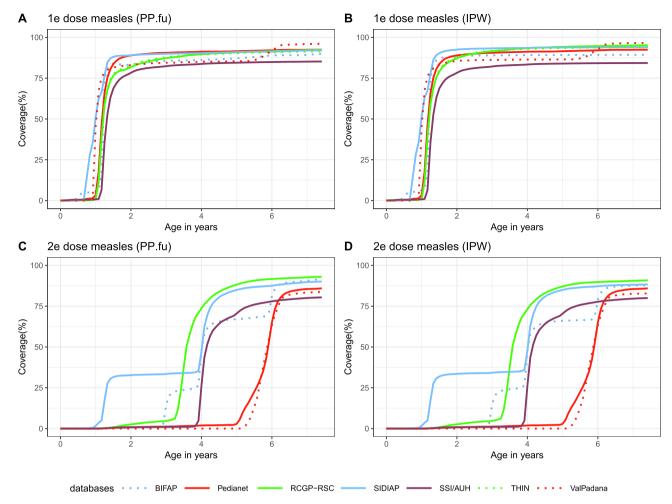


Fig. 2. Coverage with measles-vaccine by age. A: first dose, PP.fu-estimate, B: first dose, IPW-estimate, C: second dose, PP.fu-estimate, D: second dose, IPW-estimate over age in years by database for measles-vaccine, birth year 2006.

Table 3Measles coverage estimates for the first and second dose at seven years of age and bootstrapped 95% confidence interval.

Country	Database	First dose measles at age	7 years	Second dose measles at a	ge 7 years
		PP.fu-estimate	IPW-estimate	PP.fu-estimate	IPW-estimate
Italy	Val Padana	96.1% (95.6–97.2%)	96.5% (95.8–97.9%)	83.8% (83.3–85.3%)	82.8% (82.3-84.3%)
•	Pedianet	92.4% (92-93.2%)	92.4% (92-93.2%)	86% (85.3-87%)	86% (85.3-87%)
Spain	BIFAP	90% (89.5-90.5%)	89.3% (88.9-89.5%)	91.9% (91.6-92.1%)	87.8% (87.5-88%)
•	SIDIAP	92.1% (91.7-92.3%)	94% (94–95.2%)	90.1% (89.5-90.6%)	88.3% (84.6-85.4%)
UK	RCGP-RSC	92.3% (91.6–93%)	95.2% (94.5-95.9%)	93.1% (92.3-93.9%)	90.9% (90.1-91.7%)
	THIN	91.5% (91.2-91.7%)	94.1% (93.8-94.3%)	1	1
Denmark	SSI/AUH	85.2% (85–85.5%)	84.3% (84.1-84.5%)	80.4% (80.1–80.6%)	80.0% (79.7–80.2%)

Table 4 HPV-vaccine, first dose coverage estimates (IPW & PP.fu-method) at the age of 15 years aggregated over birth years 1997–2000, females.

	Country	Database	PP.fu-estimate	IPW-estimate
Ī	Italy	Val Padana	73.7%	75.8%
	Spain	BIFAP	78.8%	86.9%
		SIDIAP	77.5%	77.0%
	UK	RCGP-RSC	56.2%	60.0%
		THIN	57.8%	60.0%
	Denmark	SSI/AUH	88.6%	88.3%

reasons for entering or leaving follow-up therefore would be interesting topics for further research.

The IPW-method accounts for both left and right censoring of vaccinations, but can produce unstable estimates when weights

are very small or large and bias can accumulate as the method sums over the weekly estimated number of vaccinations. In the BIFAP-database the large difference between the IPW (86.9%) and PP.fu (78.8%)-estimates for the first HPV-vaccine dose was likely caused by the high proportion of incomplete follow-up (<50%) over a longer time period. In such instances the PP.fu-estimates might be preferred.

We recommend to always present and describe the incompleteness of the databases and to apply several methods. Large differences between methods can indicate left censoring, unstable estimation because of little data or a violation of the assumption of an equal age-specific probability of vaccination in and out of follow-up. An unequal vaccination probability in and out of follow-up will bias both methods, but it will result in a bias that

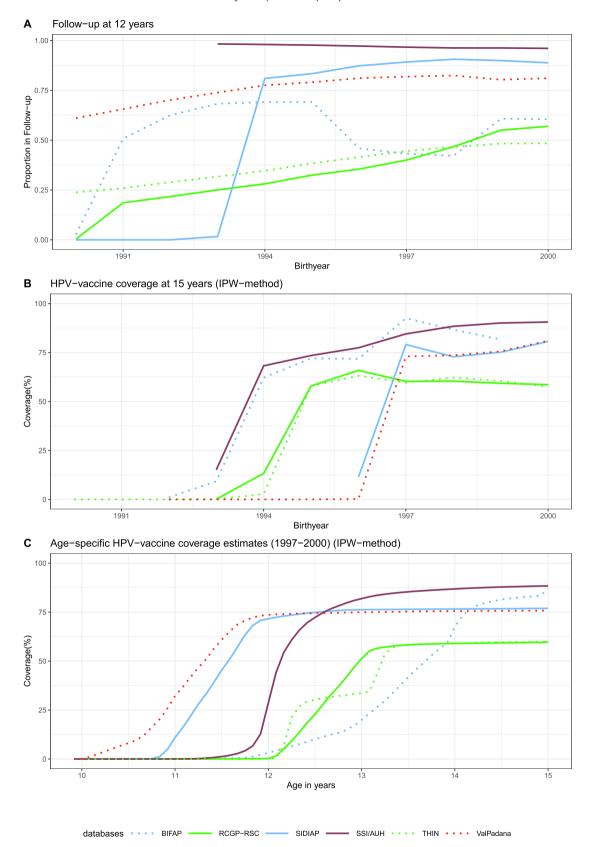


Fig. 3. A: proportion of females in follow-up at the age of 12 years over birth years by database, B: HPV-vaccine, first dose coverage estimates at 15 years (IPW-method) over birth years (1990–2000) by database, C: age-specific first dose HPV-vaccine coverage estimates (IPW-method) over age by database, birth years 1997–2000.

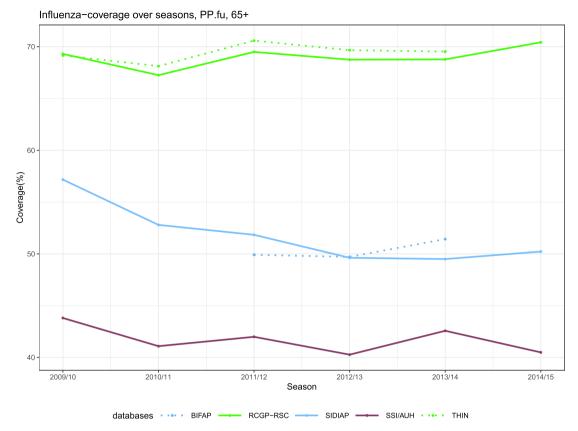


Fig. 4. Influenza-vaccine coverage estimates (PP.fu-method) for those aged over 65 years over influenza season by database, birth years 1920–1950, influenza seasons 2010–2015.

accumulates over the age-specific estimates obtained with the IPW-method (as the estimated number of vaccines at 'age in weeks' t-1,t-2,... is used for estimating the coverage at 'age in week' t), while it will bias the PP.fu-estimates for each 'age in week'-estimate independently. It is therefore likely to result in a larger bias for the IPW-method (this can be further explored in our corresponding simulation study [17]).

We only accounted for censoring because of incomplete followup. We assumed no incorrect or missing registration during followup. Previous studies have shown that the presence of a vaccination record in an electronic database reliably indicates immunization while the absence of such a record can be inaccurate [24]. This therefore was a strong assumption; e.g. vaccinations may be administered by non-traditional providers who were not affiliated with the database [4,25].

4.2. Comparison to published coverage estimates

Researchers have been concerned with the validity and comparability of coverage estimates over different studies, areas and time [35]. As methodological differences and differences in data source have proven to be relevant [3,36], they should be taken into consideration when comparing our estimates to previously published estimates. We first considered the reported outcome; our estimates were reported as age-specific estimates by birth year. Other research might report point estimates for age groups by calendar year. In addition to the heterogeneity in outcomes, we anticipated other methodological differences. Our methodology was designed specifically to account for the dynamic follow-up of populations in eHR-databases. While survival methods have been applied in coverage estimation to investigate timeliness of vaccine uptake

and right-censoring [37], accounting for both left- and right-censoring is, to our knowledge, unique to this work. Finally, differences in coverage estimation can be linked to the source of the data. For example, the databases linked to primary care from certain regions might not be representative for the national level. In addition, we added an inclusion criteria to some of our examples, possibly further limiting the comparability of our estimates to previously published estimates. Therefore, while we anticipated some differences, we also expected them to have multiple causes and it was unclear what the clinical significance (i.e., the magnitude of impact) of these differences was [36]. An individual assessment for all vaccines by database was necessary.

In recent years several measles outbreaks have been reported in the European Region associated to immunity gaps; pockets of unvaccinated children and adults [38]. From the four countries in this study, Spain was the only country with a coverage over 95% for the first dose, none obtained a coverage over 95% for the second dose. The IPW-estimates for the first dose of measles-vaccine were both below (2-8% Denmark and Spain) and above (2-9% UK and Italy) previously published national estimates (Table 5). For the second dose, our estimates were generally below (3-5%) previously published national estimates. We observed some issues with the dose of measles-vaccines in Val Padana. A sudden increase in first dose coverage sometime after the recommended age of vaccination probably reflected wrongly coding the second dose as the first. Val Padana showed the highest coverage with the first dose of measles and the lowest with the second dose. The first dose coverage increased suddenly at age 5, which is the recommended age for the second dose. The Val Padana database did not register the dose at administration but derived it afterwards from the sequence of observed vaccina-

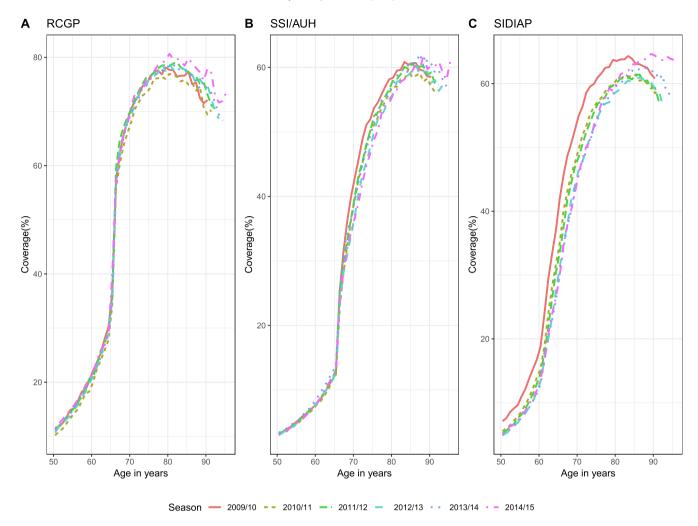


Fig. 5. Influenza-vaccine coverage (PP.fu-method) over age in years by season, (A) RCGP-RSC, (B) SSI/AUH, (C) SIDIAP, birth years 1920–1950, influenza seasons in 2009–2015.

tions. We observed differences between databases from the same country: for the measles-vaccine we observed a 1–2% difference between SIDIAP and BIFAP (Spain) and 3–4% between Pedianet and Val Padana (Italy). The differences were small given that they collected data from different regions. In 2006 an unprecedented outbreak occurred in Catalonia, Spain, affecting mostly young children (<16 months) [39]. This resulted in an early administration of the first dose of measles vaccine, before the age of one year, followed by an early administration of the second measles dose, by the age of 12–15 months. This explained the 30% second dose coverage at around 15 months followed by the, anticipated, second increase in coverage at 4 years.

Since 2007 (HPV-vaccine was licensed in 2006), HPV-vaccination has been implemented in Western-Europe. The targeted population typically consists of girls aged 10–14 years. Several modelling studies have suggested coverage levels at which HPV serotypes can be eradicated ranging from 30 to 66% to 86–94% depending on the serotype [40]. In 2014, the overall coverage of the first dose of HPV in the European region was estimated at 49.6% (95% CI 40–54%) for females aged 10–20 [12]. The four countries included in this study had previously reported high coverages (74–91%) [26–28], but estimates are known to vary over different birth cohorts in the same country. For example, in Denmark a significant decrease has been reported: a first dose coverage of 81% was reported for girls born in 2002, the coverage decreased to 54% for girls born in 2003 and continued to decrease for later birth years [30]. Our estimates are in general in agreement (+-5%) with

previously published estimates except for the UK estimates. Official UK HPV-vaccine coverage estimates for the first dose have been around 85% since 2006. We however found a lower estimate of 60% that was consistent over time for both databases. The most likely explanation is that HPV-vaccination is not always provided by the GP, since it is routinely offered in schools to girls aged 12 to 13 years. The BIFAP database recently conducted a study to validate their HPV-vaccination data quality and confirmed that BIFAP was 'a potential data source for HPV vaccine research' [41]. They reported region-specific coverage between 70.6 and 99.8% for birth cohorts born between 2000 and 2002, which included the 78.8% obtained in this study [42].

Previous research reported that influenza-vaccine coverage has been low in persons aged 65 years and more. Only a few countries, among which the UK, had achieved 75% coverage in the period 2009–2011 [33]. Influenza-vaccine coverage was estimated at 58.6% in the 2010–2011 season and 62.7% in the 2009–2010 season in persons aged 65 years and more in Navarre, Spain. In the 2012–2013 season the coverage in Madrid was estimated at 56.6% in persons aged 65 years and more [43]. Our influenza estimates by season were below (2–7%) these previously published estimates for all databases, but showed comparable time trends. Our results on influenza coverage by age agreed with previous research and presented the age-group from 77 to 94 years as the age-group with the highest coverage.

There are several limitations associated with our approach. First the creation of the input-files was a database-specific process. The

dose and country compared to the estimates obtained in this study (IPW-estimates for measles-vaccine because of left-censoring, IP. fu-estimates for HPV-vaccine because of the If estimates were only procertain birth years and PP. fu estimates for influenza-vaccine). If estimates were only available for a different birth year this is specified in bold. Coverage of measles-vaccines were obtained from WHO For the other vaccines the reference is reported next to the estimates. "(Since WHO does not report coverage by birth year, the first dose measles estimates were obtained for calendar year 2007, second dose estimates for calendar rear 2009), '(For SSI/AUH the published national estimates relied on the same data source as used for this study) Published national vaccination coverage by vaccine type,

Birth year	Italy			UK			Spain			Denmark	
	National Pedian 1e Dose Measles	Pedianet aasles	National Pedianet Val Padana National 1e Dose Measles	National	RCGP-RSC THIN	THIN	National	BIFAP	BIFAP SIDIAP	National	SSI/AUH
*5006	88	92.4	96.5	98	95.2	94.1			94		84.3
*5006	84	98	82.6	79	6.06	ı	06	87.8	88.3	85+	80
	1e Dose HPV	Λc									
2000	74.1	ı	73.7	91 [26]	56.2	57.8	77.8 [27]	78.8	77.5	90 ⁺ [29]	88.6
	(1997)						87.2 (1996) [28]			90-92* [30]	
Influenza seasons Influenza in persons aged 65 years and more	Influenza in	1 persons age	ed 65 years and	1 more							
2009–2015		. 1	ı	75% [31]	67-71	67-71		52-50	57-50	52-50 57-50 Decreasing 50->40%* (2008-2015) [33] 44-40	44-40
				71% (2016–17) [32]			(drop after 2009) 65->55% (2008-11) [33]				

process from original data to data in the common data model, included steps of data cleaning, handling missing data, removing duplicates, which were imperfectly documented and likely differed between databases. As a consequence, some of the differences within and between databases (e.g. changes in data collection over time) and with published national estimates could not be fully explored. Future studies should further explore these differences. These studies will require participation by the data access provider (input-file creation and access) and a thorough understanding of the R-scripts used for data transformation and analysis. In addition, there should also be a detailed insight into how the reference estimates were calculated.

Second, closely related to the first point, dose derivation in the absence of a recorded dose, would ideally be part of the common procedure. Given the current absence of guidelines for deriving a dose number, it is unclear how best to approach this. Third, in this paper the exact vaccines are not specified. Influenza-vaccine might for example represent a trivalent or a quadrivalent vaccine. The within-database coding of such vaccine specifications might differ between databases. Finally, we did not look into coverage with a sequence of vaccines. We reported coverage with the first and coverage with the second dose of measles-vaccine, but not coverage with both doses. Likewise for the influenza-vaccine we could not identify the proportion of persons vaccinated during, for example, both the influenza seasons 2010/11 and 2011/12.

5. Conclusion

As part of a fit for purpose assessment we estimated the vaccination coverage for measles, HPV and influenza-vaccines in seven European eHR databases. The distributed approach, with a common data model and common analysis, was feasible for all participating databases. The similar way in which the age-specific coverage is estimated, facilitates further use in international vaccine studies. The comparison with published reference coverage estimates showed validity of our procedure and the reliability of eHR-databases as source for coverage estimation, but an individual database-specific assessment of the results remains essential.

CRediT authorship contribution statement

Toon Braeye: Methodology, Software, Validation, Formal analysis, Visualization. Hanne-Dorthe Emborg: Formal analysis, Writing - review & editing, Data curation. Ana Llorente-García: Formal analysis, Writing - review & editing, Data curation. Consuelo Huerta: Formal analysis, Writing - review & editing, Data curation. Elisa Martn-Merino: Formal analysis, Writing - review & editing, Data curation. Talita Duarte-Salles: Formal analysis, Writing - review & editing, Data curation. Giorgia Danieli: Formal analysis, Writing - review & editing, Data curation. Lara Tramontan: Formal analysis, Writing - review & editing, Data curation. Daniel Weibel: Formal analysis, Writing - review & editing, Data curation. Chris McGee: Formal analysis, Writing - review & editing, Data curation. Marco Villa: Formal analysis, Writing - review & editing, Data curation. Rosa Gini: Formal analysis, Writing - review & editing, Data curation. Matti Lehtinen: Writing - review & editing. Lina Titievsky: Supervision, Project administration. Miriam Sturkenboom: Methodology, Supervision, Project administration.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Toon Braeye, Hanne-Dorthe Emborg, Ana Llorente-García, Consuelo Huerta, Elisa Martín-Merino, Talita Duarte-Salles,

Giorgia Danieli, Lara Tramontan, Daniel Weibel, Chris McGee, Marco Villa, Matti Lehtinen declared no potential conflicts of interest. Miriam Sturkenboom, declared that she has received grants from Novartis, CDC and Bill & Melinda Gates Foundation, for work unrelated to the submitted work. Lina Titievsky declared that she is employed by Pfizer and holds company stocks/shares. Rosa Gini's institution is involved in studies funded by Novartis, Eli Lilly, and Daiichi, compliant with the ENCEPP CoC.

Acknowledgements

The Innovative Medicines Initiative Joint Undertaking funded this project under ADVANCE grant agreement n° 115557, resources of which were composed of a financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and in kind contributions from EFPIA member companies.

Declaration on authorship

All authors have contributed to and approved the final article.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.vaccine.2020.02.082.

References

- [1] Greenberg M, Simondon F, Saadatian-Elahi M. Perspectives on benefit-risk decision-making in vaccinology: conference report. Hum Vaccines Immunother 2015;12:176–81. https://doi.org/10.1080/21645515.2015.1075679.
- [2] Baker MA, Nguyen M, Cole DV, Lee GM, Lieu TA. Post-licensure rapid immunization safety monitoring program (PRISM) data characterization. Vaccine 2013;31(Suppl 10):K98-K112. https://doi.org/10.1016/j.jvaccine.2013.04.088.
- [3] Miles M, Ryman TK, Dietz V, Zell E, Luman ET. Validity of vaccination cards and parental recall to estimate vaccination coverage: a systematic review of the literature. Vaccine 2013;31:1560–8. https://doi.org/10.1016/j.vaccine.2012.10.089.
- [4] Jiménez-García R, Hernandez-Barrera V, Rodríguez-Rieiro C, Carrasco Garrido P, López de Andres A, Jimenez-Trujillo I, et al. Comparison of self-report influenza vaccination coverage with data from a population based computerized vaccination registry and factors associated with discordance. Vaccine 2014;32:4386-92. https://doi.org/10.1016/j.vaccine.2014.06.074.
- [5] Khare M, Piccinino L, Barker LE, Linkins RW. Assessment of immunization registry databases as supplemental sources of data to improve ascertainment of vaccination coverage estimates in the national immunization survey. Arch Pediatr Adolesc Med 2006;160:838–42. https://doi.org/10.1001/archpedi.160.8.838.
- [6] Trinh L, Macartney K, McIntyre P, Chiu C, Dey A, Menzies R. Investigating adverse events following immunisation with pneumococcal polysaccharide vaccine using electronic General Practice data. Vaccine 2017;35:1524–9. https://doi.org/10.1016/j.vaccine.2017.01.063.
- [7] Regan AK, Mak DB, Moore HC, Tracey L, Saker R, Jones C, et al. Surveillance of antenatal influenza vaccination: validity of current systems and recommendations for improvement. BMC Public Health 2015;15:1155. https://doi.org/10.1186/s12889-015-2234-z.
- [8] Samuels RC, Appel L, Reddy SI, Tilson RS. Improving accuracy in a computerized immunization registry. Ambul Pediatr Off J Ambul Pediatr Assoc 2002;2:187–92.
- [9] MacDonald SE, Schopflocher DP, Golonka RP. The pot calling the kettle black: the extent and type of errors in a computerized immunization registry and by parent report. BMC Pediatr 2014;14:1. https://doi.org/10.1186/1471-2431-14-
- [10] Lopalco PL, Santisteve PC. Actual immunization coverage throughout Europe: are existing data sufficient?. Clin Microbiol Infect 2014;20:7–11. https://doi.org/10.1111/1469-0691.12453.
- [11] Haverkate M, D'Ancona F, Johansen K, van der Velden K, Giesecke J, Lopalco PL. Assessing vaccination coverage in the European Union: is it still a challenge?. Expert Rev Vaccines 2011;10:1195–205. https://doi.org/10.1586/erv.11.87.
- [12] Bruni L, Diaz M, Barrionuevo-Rosas L, Herrero R, Bray F, Bosch FX, et al. Global estimates of human papillomavirus vaccination coverage by region and income level: a pooled analysis. Lancet Glob Health 2016;4:e453–63. https://doi.org/10.1016/S2214-109X(16)30099-7.
- [13] Diseases CIS for I. CISID The Computerized Information System for Infectious Diseases n.d. http://data.euro.who.int/cisid/ (accessed February 20, 2020).

- [14] VENICE III n.d. http://venice.cineca.org/the_project.html (accessed August 10, 2018).
- [15] Blank PR, Schwenkglenks M, Szucs TD. Vaccination coverage rates in eleven European countries during two consecutive influenza seasons. J Infect 2009;58:446–58. https://doi.org/10.1016/j.jinf.2009.04.001.
- [16] Blank PR, Schwenkglenks M, Szucs TD. Influenza vaccination coverage rates in five European countries during season 2006/07 and trends over six consecutive seasons. BMC Public Health 2008;8:272. https://doi.org/10.1186/ 1471-7458-8-772
- [17] Braeye T, Bauchau V, Sturkenboom M, Emborg H-D, García AL, Huerta C, et al. Estimation of vaccination coverage from electronic healthcare records; methods performance evaluation – A contribution of the ADVANCE-project. PLoS ONE 2019;14:e0222296. https://doi.org/10.1371/journal.pone.0222296.
- [18] Sturkenboom M. Advancing collaborative vaccine benefits and safety research in Europe via the ADVANCE code of conduct. Vaccine 2018;36:194–5. https://doi.org/10.1016/j.vaccine.2017.08.043.
- [19] Sturkenboom M, Braeye T, van der Aa L, Danieli G, Dodd C, Duarte-Salles T, et al. ADVANCE database characterisation and fit for purpose assessment for multi-country studies on the coverage, benefits and risks of pertussis vaccinations. Vaccine 2020. https://doi.org/10.1016/j.vaccine.2020.01.100.
- [20] Vaccine Scheduler | ECDC n.d. https://vaccine-schedule.ecdc.europa.eu/ (accessed August 10, 2018).
- [21] Walton S, Cortina-Borja M, Dezateux C, Griffiths LJ, Tingay K, Akbari A, et al. Measuring the timeliness of childhood vaccinations: Using cohort data and routine health records to evaluate quality of immunisation services. Vaccine 2017;35:7166-73. https://doi.org/10.1016/j.vaccine.2017.10.085.
- [22] Cain KC, Harlow SD, Little RJ, Nan B, Yosef M, Taffe JR, et al. Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. Am J Epidemiol 2011;173:1078–84. https://doi.org/ 10.1093/aie/kwo481.
- [23] Hazelbag CM, Klungel OH, van Staa TP, de Boer A, Groenwold RHH. Left truncation results in substantial bias of the relation between time-dependent exposures and adverse events. Ann Epidemiol 2015;25:590–6. https://doi.org/10.1016/j.annepidem.2015.03.019.
- [24] Sy LS, Liu I-LA, Solano Z, Cheetham TC, Lugg MM, Greene SK, et al. Accuracy of influenza vaccination status in a computer-based immunization tracking system of a managed care organization. Vaccine 2010;28:5254–9. https://doi. org/10.1016/j.vaccine.2010.05.061.
- [25] Irving SA, Donahue JG, Shay DK, Ellis-Coyle TL, Belongia EA. Evaluation of self-reported and registry-based influenza vaccination status in a Wisconsin cohort. Vaccine 2009;27:6546–9. https://doi.org/10.1016/j.vaccine.2009.08.050.
- [26] Annual HPV vaccine coverage 2013 to 2014: by PCT, local authority and area team. GOVUK n.d. https://www.gov.uk/government/statistics/annual-hpvvaccine-coverage-2013-to-2014-by-pct-local-authority-and-area-team (accessed August 11, 2018).
- [27] Ministerio de Sanidad, Consumo y Bienestar Social Profesionales Vacunas Coberturas de Vacunación n.d. http://www.mspsi.gob.es/profesionales/ saludPublica/prevPromocion/vacunaciones/coberturas.htm#primero (accessed August 11, 2018).
- [28] Limia A, Pachón I. Coverage of human papillomavirus vaccination during the first year of its introduction in Spain. Eurosurveillance 2011;16:19873. https://doi.org/10.2807/ese.16.21.19873-en.
- [29] Skorstengaard M, Thamsborg LH, Lynge E. Burden of HPV-caused cancers in Denmark and the potential effect of HPV-vaccination. Vaccine 2017;35:5939-45. https://doi.org/10.1016/j.vaccine.2017.08.062.
- [30] Suppli CH, Hansen ND, Rasmussen M, Valentiner-Branth P, Krause TG, Mølbak K. Decline in HPV-vaccination uptake in Denmark the association between HPV-related media coverage and HPV-vaccination. BMC Public Health 2018;18:1360. https://doi.org/10.1186/s12889-018-6268-x.
- [31] Vaccine uptake guidance and the latest coverage data. GOVUK n.d. https://www.gov.uk/government/collections/vaccine-uptake (accessed August 11, 2018).
- [32] Seasonal flu vaccine uptake in GP patients in England: winter season 2016 to 2017. GOVUK n.d. https://www.gov.uk/government/statistics/seasonal-fluvaccine-uptake-in-gp-patients-in-england-winter-season-2016-to-2017 (accessed August 11, 2018).
- [33] Mereckiene J, Cotter S, Nicoll A, Lopalco P, Noori T, Weber JT, et al. Seasonal influenza immunisation in Europe. Overview of recommendations and vaccination coverage for three seasons: pre-pandemic (2008/09), pandemic (2009/10) and post-pandemic (2010/11). Eurosurveillance 2014;19:20780. https://doi.org/10.2807/1560-7917.ES2014.19.16.20780.
- [34] WHO World Health Organization: Immunization, Vaccines And Biologicals. Vaccine preventable diseases Vaccines monitoring system 2018 Global Summary Reference Time Series: MCV1 n.d. http://apps.who.int/ immunization_monitoring/globalsummary/timeseries/tscoveragemcv1.html (accessed August 11, 2018).
- [35] Murray CJ, Shengelia B, Gupta N, Moussavi S, Tandon A, Thieren M. Validity of reported vaccination coverage in 45 countries. The Lancet 2003;362:1022-7. https://doi.org/10.1016/S0140-6736(03)14411-X.
- [36] Wilson SE, Quach S, MacDonald SE, Naus M, Deeks SL, Crowcroft NS, et al. Methods used for immunization coverage assessment in Canada, a Canadian immunization research network (CIRN) study. Hum Vaccines Immunother 2017;13:1928–36. https://doi.org/10.1080/21645515.2017.1319022.
- [37] Laubereau B, Hermann M, Schmitt HJ, Weil J, Kries RV. Detection of delayed vaccinations: a new approach to visualize vaccine uptake. Epidemiol Amp Infect 2002;128:185–92. https://doi.org/10.1017/S0950268801006550.

- [38] Datta SS, O'Connor PM, Jankovic D, Muscat M, Ben Mamou MC, Singh S, et al. Progress and challenges in measles and rubella elimination in the WHO European Region. Vaccine 2018;36:5408–15. https://doi.org/10.1016/j.vaccine.2017.06.042.
- [39] Torner N, Martinez A, Costa J, Mosquera MM, Barrabeig I, Rovira A, et al. Measles outbreak in the Barcelona Region of Catalonia, Spain, October 2006 to February 2007. Wkly Releases 1997–2007 2007;12:3144. https://doi.org/ 10.2807/esw.12.08.03144-en.
- [40] Lehtinen M, Baussano I, Paavonen J, Vänskä S, Dillner J. Eradication of human papillomavirus and elimination of HPV-related diseases - scientific basis for global public health policies. Expert Rev Vaccines 2019;18:153–60. https://doi. org/10.1080/14760584.2019.1568876.
- [41] Martín-Merino E, Llorente-García A, Montero-Corominas D, Huerta C. The recording of human papillomavirus (HPV) vaccination in BIFAP primary care database: a validation study. Pharmacoepidemiol Drug Saf 2019;28:201–8. https://doi.org/10.1002/pds.4674.
- [42] Martín-Merino E, Llorente-García A, Montero Corominas D, Huerta-Álvarez C. The longitudinal incidence of human papillomavirus vaccination in spanish primary care in the first 6 years after approval. Pharm Med 2019. https://doi.org/10.1007/s40290-019-00268-1.
- [43] Jiménez-García R, Esteban-Vasallo MD, Rodríguez-Rieiro C, Hernandez-Barrera V, Domínguez-Berjón MF, Garrido PC, et al. Coverage and predictors of vaccination against 2012/13 seasonal influenza in Madrid. Spain. Hum Vaccines Immunother 2014;10:449–55. https://doi.org/10.4161/hv.27152.