BMJ Open

Effectiveness of HPV vaccine by age at vaccination and number of doses: protocol for a population-based matched case–control study

Carlos R Oliveira,[1] Anette Michelle Ortiz ,[1] Sangini S Sheth,[2] Eugene D Shapiro,[1,3] Linda M Niccolai[3]

To cite: Oliveira CR, Ortiz AM, Sheth SS, *et al*. Effectiveness of HPV vaccine by age at vaccination and number of doses: protocol for a population-based matched case–control study. *BMJ Open* 2021;**11**:e043093. doi:10.1136/bmjopen-2020-043093

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2020-043093).

Received 30 July 2020
Revised 31 January 2021
Accepted 07 April 2021

[1]Pediatrics, Yale University School of Medicine, New Haven, Connecticut, USA
[2]Obstetrics, Gynecology, & Reproductive Sciences, Yale School of Medicine, New Haven, Connecticut, USA
[3]Epidemiology of Microbial Diseases, Yale University School of Public Health, New Haven, Connecticut, USA

Correspondence to
Dr Carlos R Oliveira;
carlos.oliveira@yale.edu

ABSTRACT

Introduction In 2006, the first human papillomavirus (HPV) vaccine was approved by the Food and Drug Administration in the USA based on pre-licensure clinical trials that found it to be highly efficacious at preventing persistent infection and precancerous, high-grade cervical lesions (HGCLs) caused by viral types the vaccine protects against. However, the real-world effectiveness of HPV vaccines as used in clinical practice may be quite different from the efficacy found in pre-licensure clinical trials. More than 10 years have passed since the introduction of the vaccine programme. It is critical to determine if the full benefits of HPV are being realised in real-world settings.

Methods and analysis The objectives of this study were to estimate the effectiveness of HPV vaccines as used in real-world clinical settings and to determine the degree to which the vaccine's effectiveness varies based on age at the time of immunisation and the number of doses received. The study will be a population-based, matched case–control study. Cases will be women with newly diagnosed HGCL associated with HPV types 16 and 18. Matched controls will be women with a normal Pap test result, matched individually to cases in a 2:1 ratio by age, a practice and date of testing. Medical records will be reviewed to determine dates of receipt of the HPV vaccine for all participants. We will use multivariate conditional logistic regression to control for potential confounders.

Ethics and dissemination This protocol presents minimal risk to the subjects. This protocol has received approval from the Institutional Review Board of Yale University (HIC: 1502015308), and a Health Insurance Portability and Accountability Act (HIPAA) Waiver of Authorisation has been granted to allow investigators to recruit subjects for the study. Findings will be disseminated through peer-reviewed, open-access scientific journals and conference presentations.

Strengths and limitations of this study

- Population-based case–control design is an efficient means of determining the vaccine's effectiveness (VE) after licensure.
- Comprehensive collection of data on immunisation history and potential confounders will produce the most valid estimates of effectiveness by age at the time of vaccination and the number of doses received.
- Data will provide empirical estimates of the VE when given at different ages.
- Case–control studies are potentially subject to important biases including selection and information bias.
- Given the long latent period from human papillomavirus infection to invasive cervical carcinoma, the timing of this study does not permit estimates of the VE against cervical cancer.

INTRODUCTION

Human papillomavirus (HPV) is the most common sexually transmitted infection in the USA.[1] There are over 100 types of HPV, approximately 15 of which are considered oncogenic (high-risk) types for the development of cervical cancer.[2] The US Food and Drug Administration has licensed three vaccines to prevent infection with HPV.[3–5] All licensed HPV vaccines were evaluated through pre-licensure randomised clinical trials, which found high efficacy (97%–98%) against precancerous high-grade cervical lesions (HGCLs) for women not previously infected with the types of HPV in the vaccine.[6 7] The proposed work aims to address several important questions that remain about the extent to which HPV vaccines are realising their full potential for preventing HGCL.

First, what is the real-world effectiveness of the vaccine? A vaccine's *efficacy* is measured pre-licensure, through a clinical trial and in highly controlled research settings. The goal of a vaccine efficacy trial is to estimate the maximum potential benefit of a vaccine (ie, the protective effect of the vaccine in an ideal scenario). In contrast, studies of a vaccine's *effectiveness* measure the vaccine's protective effect post-licensure and in real-world clinical settings. The protective effect of a vaccine in the post-licensure clinical settings (ie, its effectiveness) does not always equate with the benefits anticipated from the pre-licensure clinical trials (ie, the efficacy). Thus, post-licensure studies that assess the vaccine's effectiveness (VE) are crucial. Few studies have been published that have examined the real-world, individual-level effectiveness of the HPV vaccine against HGCL.[8–11] Although these studies represent an important first step in documenting the effectiveness of HPV vaccines, they have had several important limitations. For instance, these studies relied on disease registries with limited individual-level information and, thus, were unable to adequately control for potential confounders in their analyses. Furthermore, most of the prior work did not have information on the types of HPV, so they could not estimate the type-specific effectiveness of the HPV vaccine.

Second, does the VE vary by age at the time of immunisation? Even with evidence of high efficacy in pre-licensure stages, HPV vaccine coverage remains suboptimal in the USA, with only 54.2% of adolescent females and 32.8% of males aged 13–17 years having completed the series.[12] This coverage is substantially lower than that seen in other industrialised countries, including Australia, Denmark and England, where >70% of adolescent girls have completed the series.[13–15] It is also lower than the rates of immunisation in the USA for other vaccines routinely given to adolescents.[12] Among adolescents, only 39.0% have shown to have completed the entire vaccine series by 13 years of age. Prior studies have suggested that HPV VE may decline with increasing age at the time of immunisation[16]; however, empiric data on the effectiveness of the vaccine at younger ages are still unknown.

Third, what is the effectiveness of a two-dose schedule? The Advisory Committee on Immunisation Practices (ACIP) recommends that a two-dose immunisation regimen be used if the first dose can be given prior to age 15 years. The ACIP recommends a three-dose series if the first dose is given after age 15 years.[17] Achieving adequate coverage with two or three doses is a challenge because it requires additional healthcare visits beyond recommended annual preventive visits. Numerous barriers to completing the vaccine series have been identified and include low awareness of the need for additional doses. These challenges result in low completion rates and lack of a strong recommendation from a healthcare provider. However, early reports indicate that the immunogenicity of two and three doses of the HPV vaccine may be similar,[18–21] which raised the interest in assessing the clinical effectiveness of a two-dose regimen more broadly.[22] A two-dose regimen for all age groups would reduce logistical and resource challenges and may be of particular value in settings where access to the vaccine is limited.

To answer these important questions, we designed a matched case–control study to determine the effectiveness of HPV vaccines against incident HGCL attributable to HPV 16 or 18 (oncogenic HPV types in available vaccines) and to estimate the extent to which age at the time of immunisation and number of doses received influences the vaccine's real-world effectiveness.

METHODS AND ANALYSIS

Overview

To estimate the effectiveness of HPV vaccines in real-world settings, we propose a population-based matched case–control study. For this study, case subjects will be women with an HGCL (cervical intraepithelial neoplasia grades 2 or 3 or adenocarcinoma in situ) that tested positive for HPV 16 or 18. These women will be identified from a statewide registry in Connecticut that was established in 2008 to monitor trends in HGCL after introduction of HPV vaccines. This registry that gathers information on women diagnosed with HGCL has been used to report trends in incidence of disease since vaccine introduction and also provides the source population for cases in this study.[23 24] Matched controls will be women with a normal Pap test result, who are individually matched to cases by age, medical practice and focal time. Focal time will be defined for cases as the date of the 'trigger Pap', that is, the date of the abnormal Pap test that preceded the histopathological diagnosis of HGCL. For controls, the focal time will be defined as the date of their normal Pap test. The source population for the controls will be female residents of New Haven County (population: 860 435), born in 1981 or later, whose cervical cytology samples obtained for screening were analysed by the Yale Department of Pathology during the study period. Eligible cases will be obtained through collaboration with a surveillance programme registry known as the 'HPV Vaccine Impact Monitoring Project Across CT' (HPV-IMPACT). The birth year criterion will ensure that all cases and controls will have been eligible for the HPV vaccine by having been 26 years of age or younger in 2006 when vaccinations were first approved for this age group. Including women whose cervical screening test was processed at a single pathology laboratory will provide an accessible sampling frame for selecting controls and will allow us to focus on a select number of medical practices for the efficiency of data collection, while still achieving adequate sample size. Additionally, a single pathology laboratory would decrease variability in cytology and histopathology diagnosis. All participants will be asked to complete a survey that will allow us to evaluate for potential confounders and to obtain information on immunisation history. To verify self-reported immunisation history, medical records will be reviewed from all reported prior sources of care. An overview of our study design and case definitions are presented in figure 1.

Several study designs have been used to quantify the real-world effectiveness of vaccines.[25–27] In a clinical trial, randomisation serves to decrease the risk of introducing bias in the allocation of the vaccine and/or confounding by differences in treatment groups. However, after a vaccine has been shown to be efficacious and has been approved for routine use, it becomes unethical to randomise subjects to receive a placebo vaccine. Hence, observational studies are the most appropriate, and ethical, method for ascertaining the post-licensure effectiveness of vaccines. The two most frequently used observational methods are the cohort and case–control designs.[28] The logic of a cohort study is similar to that of a clinical trial, in which subjects are classified according to their self-determined

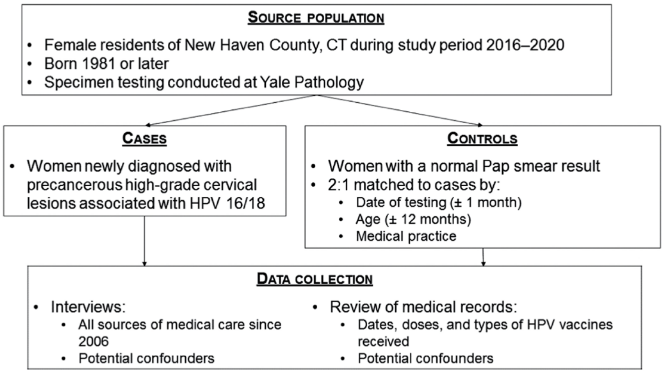

Figure 1 Study design and case–control definitions. CT, Connecticut; HPV, human papillomavirus.

(eg, not randomised) exposure status (immunised or not immunised) and followed longitudinally over a specified period of time or until disease occurs. However, in situations where there is prolonged latency between exposure and disease, as is the case with HGCL, the cohort design becomes unfeasible to conduct due to the long duration of the study. There are also ethical considerations involved in actively following a group of women known to have not received a recommended vaccine. An alternative approach to measure VE is through a case–control study. The case–control design involves selecting individuals not based on exposure but based on their outcome. In this approach, diseased patients are categorised as 'cases,' and those without the disease are 'controls.' Both cases and controls are selected from the same population source, and exposure status (immunisation history) is ascertained retrospectively (figure 2). Modern epidemiological thinking supports that the case–control design is an efficient method of sampling from an underlying cohort and thus has many of the same strengths of the cohort design but requires fewer resources to conduct.

Like all observational methods, the case–control design is susceptible to the effects of confounding. In the context of a VE study, confounding occurs when the association between immunisation and disease status is incorrect due to the effect of a third variable. Confounding variables are those that are independently associated with both the disease and the immunisation status. For example, in the case of HPV vaccine, women of lower socioeconomic status may be less likely to have access to immunisations

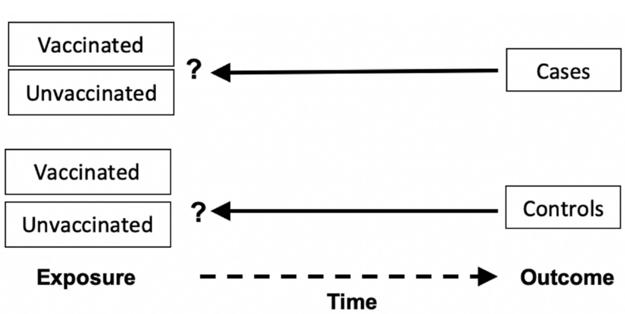

Figure 2 Case–control study design.

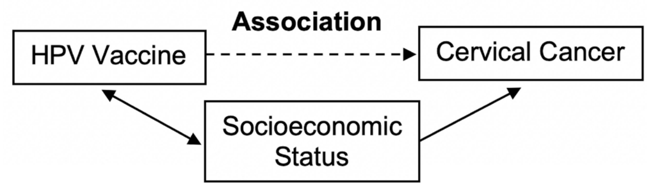

Figure 3 Socioeconomic status as a confounder for disease status. HPV, human papillomavirus.

and may also be more likely to develop cervical cancer; hence, low socioeconomic status may be a negative confounder as it is positively related to disease and negatively related to exposure (figure 3).

There are several strategies that can be employed to mitigate the effects of confounding. When confounders are known in the design phase of the study, an investigator can choose to restrict the study sample or perform individual matching of controls on the confounding variable. Restriction works by imposing homogeneity in both cases and controls with respect to the suspected confounder (ie, if socioeconomic status is a suspected confounder, you can opt to only select participants of a particular socioeconomic status). The population from which controls are selected and how closely controls resemble the cases can also drastically affect the results of the study due to bias. Sampling controls from the same catchment population (or comparable population) that the cases came from is a strategy that can be used to minimise the risk of control selection bias. Matching is another commonly used strategy to minimise the risk of control selection bias. The goal of matching is to select controls that are similar to their respective cases in all aspects relevant to the disease, with the exception of having the disease.

For the proposed HPV VE study, we will identify and enrol two controls for each enrolled case. For the selection of controls, we will use the 'test-negative' method. In this approach, patients seeking routine health maintenance are screened for cervical cancer and those testing positive (abnormal cells on Pap test) are cases, and those who test negative (normal Pap test) are controls. Hence, the 'test-negative' method limits bias due to misclassification of disease (controls cannot be an undiagnosed case) and to healthcare-seeking behaviour (both cases and controls undergoing routine health maintenance). To provide an additional degree of comparability between cases and controls, we will also incorporate matching by the date of the Pap smear, age and medical practice. Matching on the date of the test (±1 month) will control any secular trends. Matching on age (±1 year) will control for opportunity time to have received the HPV vaccine, and it will partially control for potential exposure to HPV. We will incrementally increase matching windows by 1 month for the date of the test and 1 year for age as needed to identify the sufficient number of potentially eligible controls. Matching on practice will control for confounding by practice-associated factors and will ensure similar access to medical records for both cases and controls.

We will obtain a list of potential controls and use a random number generator to establish the order in which controls will be contacted until at least two controls per case have been successfully enrolled.

The validity of a case–control VE study also depends heavily on the specificity of the case definition. The ideal approach for establishing a case definition is to select an outcome that is pathogen-specific. Case definitions that rely solely on a broad clinical diagnosis (ie, HGCL) can be useful if the pathogen being targeted by the vaccine is responsible for the majority of cases of the disease. The proportion of cases of HGCL that are attributable to an HPV strain included in the quadrivalent vaccine (types 16 and 18) is approximately 50%.[29] Although this proportion is likely decreasing with increasing proportions of populations being vaccinated,[30] we use a pathogen-specific case definition (ie, HGCL that is positive for HPV 16 or 18) for cases in this study to provide better estimates of the direct effectiveness of the quadrivalent vaccine in use during much of the study period. During the enrolment period for this study, newly diagnosed cases will be identified through HPV-IMPACT. HPV-IMPACT conducts population-based surveillance for HGCL throughout the state of Connecticut and surveillance of HPV in HGCL for residents of New Haven County.[31] For the enhanced surveillance, residual cervical specimens from women with HGCL in New Haven County are requested from the pathology laboratories and transported to the Division of High-Consequence Pathogens and Pathology laboratory at the Centers for Disease Control and Prevention (CDC), where genomic HPV DNA is extracted from the tissue and typed with the Linear Array HPV genotyping assay (LA, Roche Diagnostics, Indianapolis, Indiana, USA). This genotyping instrument's sensitivity and specificity in cervical specimens is 98% and 92%, respectively.[32] A key strength of the proposed study is the use of this population-based surveillance system as it will ensure a robust ascertainment of cases within our catchment area.

Data collection

After identifying all eligible cases and matched controls, we will send letters to potential subjects inviting them to participate in the study. The letter will include an 'opt-out' prestamped postcard to mail back if they do not wish to be contacted. Individuals who do not return the postcard by telephone will be contacted to receive an explanation of the purpose of the study and procedures required for participation. They will also be allowed to ask questions. Subjects who are willing to participate will be asked to complete a brief survey after signing informed consent. Then, we will review medical records at each practice named by the participants. Surveys and reviews of medical records will be done using identical methods for cases and controls. To minimise bias due to differential misclassification of immunisation status, we will place equal efforts to ascertain immunisation history in both cases and controls.

Table 1 Potential confounders that will be collected for inclusion in the analysis

Sociodemographic data	Survey: race, ethnicity, income, marital status and residential address Medical record: type of health insurance
Health behaviours including sexual activity	Survey: age at first sexual intercourse, number of lifetime sex partners, number of sex partners in past the 12 months, condom use and smoking status Medical record: history of another sexually transmitted infection, use of contraceptives, parity and gravity
Healthcare access and utilisation	Medical record: history of Pap screening, including dates and results of tests (frequency), number of physician office visits in the past year, receipt of other vaccines (MCV4, Tdap) recommended for ages 11–12 years and dates of administration
Health status	Medical record: immunocompromised status (major examples: HIV/AIDS, a recent transplant, history of immunosuppressive therapy or renal failure/dialysis)

Patient survey

All participants will be asked to complete a 48-question survey. Surveys will be adapted for each participant so that questions always refer to the participant's focal time. Surveys will gather information on potential confounders that cannot be readily obtained from medical records, such as sexual behaviours and smoking. Furthermore, we will ascertain the names of practices and/or physicians and locations for all sources of medical care since 2006. Variables that were identified from a review of the literature that we will ascertain are detailed in table 1.[33–40] Participants will be given the option to complete the survey electronically, in-person or over the telephone, using previously validated instruments.[41]

Patient and public involvement

Patients or public involvement is not applicable to this study. Neither patients nor the public were involved in the design, or conduct, or reporting or dissemination plans of our research.

Medical record reviews

Medical records will be reviewed for each participant at all sources of care reported or identified in the survey. We will collect information on dates, types and commercial names of all doses received at age 9 (the earliest HPV vaccine can be given) or later. We will also collect information on all potential confounders (table 1) and about additional sources of care. Trained staff will complete a standardised abstraction form at the medical practices.

Oliveira CR, *et al*. *BMJ Open* 2021;**11**:e043093. doi:10.1136/bmjopen-2020-043093

Statistical analyses

The first step in the analysis will be to produce an integrated and clean database. All data will be entered into a single database by a unique patient identifier. Cases and matched controls will be indicated. The use of computer-assisted data entry will help minimise data entry errors. Furthermore, all variables will be examined for missing and out-of-range values by examining frequency distributions of each variable. Consistency and logic checks will be conducted, and any errors will be corrected. Descriptive statistics will be used to characterise demographic, behavioural and clinical variables for the study population. Associations between independent variables will be examined using correlation coefficients, t-tests and χ^2 tests to assess possible collinearity. If detected, dimensionality will be reduced through principal component analysis.

Multivariate modelling to address our aims will be based on methods previously described.[42 43] The VE, defined as the proportionate reduction in the risk of disease among vaccinated participants that is attributable to the vaccine, will be calculated using matched ORs from a conditional logistic regression model with vaccination as exposure and case/control status as the outcome. Specified models and modelling strategies for each of our aims are described below. For all analyses, a type I error of 5% (two-sided) will be used to test for statistical significance.

Estimation of overall vaccine effectiveness

The primary exposure of interest will be receipt of HPV vaccine at least 2 years before the date of the Pap test that led to biopsy-confirmed HGCL for cases, or the date of the normal Pap test for controls. Given the natural history of HPV infections, and the amount of time needed to develop a HGCL, it is likely that vaccination less than 2 years before the diagnosis of HGCL would have occurred after infection and therefore would not be expected to prevent the outcome.[38 39 44] To assess the robustness of this assumption, we will conduct sensitivity analyses to examine the effect of using different time intervals (≥6 months prior, ≥1 year prior, ≥3 years prior and ≥5 years prior to focal time). Because all three currently approved vaccines prevent HPV 16 and 18, different vaccines will not be considered in primary analyses.

For the planned analyses, the log odds of disease for individuals in matched sets will be modelled using the following conditional logistic model:

$$\begin{aligned}\text{Model 1 : Log}(\text{odd}_{\text{case}}) &= \alpha_i + \beta_e(\text{HPV\_vaccination}) + \beta_1(\text{cov\_1}) \\ &+\beta_2(\text{cov\_2}) + \ldots + \beta_k(\text{cov\_k})\end{aligned}$$

where α_i is the stratum-specific constant term for each matched set, β_e is the parameter coefficient for the exposure of interest (receipt of HPV vaccine), and β_1 through β_k are the parameter coefficients for each covariate included in the model to control for confounding. Our modelling strategy will be to create models that always contain the exposure variable of interest (receipt of HPV vaccine) and additional covariates as needed. For each added covariate, we will assess the decrease in the log-likelihood statistic (ie, an improved fit of the overall model) until no other variable significantly improves the fit of the model. Variables with weak associations (p value >0.2) will be excluded from multivariate models. This procedure avoids overfitting the multivariable models and saves df to test the main predictor of interest. Using β_e from the model, the OR comparing vaccinated to unvaccinated women can be estimated as OR=exp(β_e), and the VE can be expressed as a percentage and calculated as VE = [1−OR]*100.

Estimation of effectiveness by age at the time of vaccination

We will use the following four a priori age categories: ≤12 years (recommended ages), 13–15 years (younger adolescence), 16–19 years (older adolescence) and ≥20 years (young adults). We will create four dichotomous variables for age at the time of receipt of the first dose that allows for variation in effect estimates across the categories where each variable is coded as '1' if vaccinated in that age group and coded as '0' otherwise (including unvaccinated). The following conditional logistic regression model will then be used to produce adjusted estimates of the VE for each age group:

$$\begin{aligned}\text{Model 2 : Log}(\text{odds}_{\text{case}}) &= \alpha_i + \beta_1(\text{age.at.vacc\_1}) + \beta_2(\text{age.at.vacz\_2}) \\ &+\beta_3(\text{age.at.vacc\_3}) + \beta_4(\text{age.at.vacc\_4}) \\ &+\beta_5(\text{cov\_1}) + \beta_6(\text{cov\_2}) + \ldots + \beta_j(\text{cov\_k})\end{aligned}$$

where β_1 is the parameter estimate for those vaccinated at ages ≤12 years compared with unvaccinated, β_2 is the parameter estimate for those vaccinated at ages 13–15 years compared with unvaccinated, β_3 is the parameter estimate for those vaccinated at ages 16–19 years compared with unvaccinated and β_4 is the parameter estimate for those vaccinated at ages ≥20 years compared with unvaccinated. To determine if there are significant differences between age groups (ie, to assess other pairwise comparisons), we will recode the dichotomous variables using different referent categories.

Estimation of effectiveness by the number of doses received

We will create the three dichotomous variables that allow for variation in estimates of effect across the categories where each variable is coded as '1' if received only that number of doses and coded as '0' otherwise (including unvaccinated). The following conditional logistic regression model will then be used to produce adjusted estimates of effectiveness for each dose compared with 0 dose, and interpreted similarly to the model for age at the time of vaccination described above:

$$\begin{aligned}\text{Model 3 : Log}(\text{odds}_{\text{case}}) &= \alpha_i + \beta_1(\text{dose\_1}) + \beta_2(\text{dose\_2}) + \beta_3(\text{dose\_3}) \\ &+\beta_4(\text{cov\_1}) + \ldots + \beta_j(\text{cov\_k})\end{aligned}$$

To determine if there are significant differences between three doses and one or two doses, we will recode the dichotomous variables using three doses as the referent category, and we will interpret the corresponding beta coefficients.

Table 2 Number of cases needed to detect different estimate of HPV vaccine's effectiveness

Proportion of controls vaccinated	Effectiveness of HPV vaccine			
	40%	50%	60%	70%
20%	342	201	127	84
30%	247	143	88	57
40%	205	116	70	45
50%	186	103	61	38
60%	182	99	58	34

HPV, human papillomavirus .

Sample size and statistical power

The number of cases needed to detect a range of estimates of the VE for aim 1, with α<0.05% and 80% power, is presented in table 2. We calculated sample sizes presented in table 2 for different proportions of controls that might be vaccinated based on the literature[9 11 45] and for the 2:1 ratio of controls to cases using established formulas.[46] The timeline for this study is anticipated to run from April 2017 to July 2022.

Ethics and dissemination

The proposed research is not a clinical trial or an intervention, and therefore, it presents no more than minimal risk to the subjects, and adverse events are not anticipated. Study participation will be completely informed, voluntary, confidential and non-discriminatory. Nobody will be excluded from this study on the basis of race or ethnicity. Decisions on whether or not to participate in this study will in no way affect the provision of medical care at the participating practices or anywhere else. Written consent will be obtained from voluntary participants only after staff has extensively discussed the study procedures with the potential participant. All research staff will be trained in ethical principles and guidelines for research. Research staff will also be fully trained in non-judgemental interactions and counselling and referral if participants become distressed during the course of this study.

All results that are derived from the proposed research will be made available through peer-reviewed publications in scientific journals and through international conference presentations. Participants will be assured that the information they provide will not be linked to them as individuals in any study reports, publication or presentations. This protocol has already received approval from the Institutional Review Board of Yale University (HIC: 1502015308), and a Health Insurance Portability and Accountability Act (HIPAA) Waiver of Authorisation has been granted to allow investigators to recruit subjects for the study.

DISCUSSION

HPV vaccines have been used in clinical practice in the USA since 2006, but their real-world effectiveness, likely to be influenced by suboptimal patterns of uptake, is not known. Delays in administration beyond the recommended ages of 11–12 years are likely to be reducing effectiveness, but no empirical estimates about the extent of the diminished effectiveness are available. Furthermore, the effectiveness based on the number of doses received is uncertain but important for designing optimal vaccination programmes. To address these gaps in knowledge, we propose the first study to evaluate the HPV VE using disease-free matched controls, type-specific outcomes and robust consideration of potential confounders.

This study has several notable strengths. First, the case–control design is an efficient means of rigorously determining the VE after licensure. Particularly, the study of a VE in preventing precancerous lesions which takes several years to develop after infection. A longitudinal cohort design would be extremely costly, time-consuming and highly subject to bias from losses to follow-up. Another aspect of the prospective design that renders it impractical is the ethical dilemma of actively following participants who are not vaccinated. Therefore, we believe that the case–control design is now necessary to address this research question. Second, our comprehensive collection of data on immunisation history and potential confounders is an important strength. This represents a significant advancement over previous research that used existing data sources and, therefore, did not include other important variables. This rigorous approach will produce the most valid estimates of effectiveness by age at the time of vaccination and the number of doses received.

Given the longer latent period from HPV infection to invasive cervical carcinoma, the timing of this study does not permit estimates of the VE to prevent cervical cancer. However, the infrastructure that we develop and the methods that we use can be applied in future studies for other diagnoses, including cancers and pre-cancers of other sites that are associated with HPV, such as the anus, vagina, vulva, penis and oropharynx. As evidence of vaccine impact on cervical cancer is just beginning to emerge,[47] this will be an important endeavour going forward.

Many parents and physicians are not aware that the VE is highly dependent on age at the time of vaccination, in part, because we do not currently have specific empirical estimates of the VE when administered at different ages. While the VE should be greater if it is administered at a younger age, because younger adolescents have a lower prevalence of infection with HPV, there are no data on the magnitude of this effect. Empirical estimates that quantify the increased benefit can be a potent incentive to increase the strength and consistency of healthcare provider recommendations and the acceptance by parents at the recommended ages of 11–12 years without delay. Furthermore, the empirical data about the effectiveness of fewer than three doses can inform policy about alternative dosing regimens that may be both easier to achieve and less costly. Answers to each of the research questions in this proposal are critically needed to provide evidence that can be used to develop strategies to more fully realise the potential benefits of HPV vaccines.

Contributors Study concept and design was fulfilled by ES, CO, SSS and LN. Drafting of the manuscript was fulfilled by AMO, CO and LN. Critical revision of the manuscript for important intellectual content was fulfilled by AMO, CO, SS, SSS and LN.

Funding This work was supported in part by the National Institutes of Health (NIH) grant numbers R01AI123204 (Dr Niccolai), KL2TR001862 (Dr Oliveira), K07CA230234 (Dr Sheth) and UL1TR000142 (Dr Shapiro) from the National Center for Advancing Translational Science (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH. The funders of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report. The corresponding author had full access to all of the data in this study and had final responsibility for the decision to submit for publication.

Competing interests Dr Niccolai has served as Scientific Advisor for Merck. Dr Sheth receives Gardasil 9 from Merck at no cost for research and has served as a consultant for Merck. All other authors declare no conflicts of interest.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Anette Michelle Ortiz http://orcid.org/0000-0002-7609-5689

REFERENCES

1 Satterwhite CL, Torrone E, Meites E, *et al*. Sexually transmitted infections among US women and men: prevalence and incidence estimates, 2008. *Sex Transm Dis* 2013;40:187–93.
2 Muñoz N, Bosch FX, de Sanjosé S, *et al*. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003;348:518–27.
3 Administration USFaD. *FDA licenses new vaccine for prevention of cervical cancer and other diseases caused by human papillomavirus*, 2006.
4 Administration USFaD. *FDA approves new vaccine for prevention of cervical cancer*, 2009.
5 Administration USFaD. *FDA approves Gardasil 9 for prevention of certain cancers caused by five additional types of HPV*, 2014.
6 FUTURE II Study Group. Quadrivalent vaccine against human papillomavirus to prevent high-grade cervical lesions. *N Engl J Med* 2007;356:1915–27.
7 Joura EA, Giuliano AR, Iversen O-E, *et al*. A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women. *N Engl J Med* 2015;372:711–23.
8 Gertig DM, Brotherton JML, Budd AC, *et al*. Impact of a population-based HPV vaccination program on cervical abnormalities: a data linkage study. *BMC Med* 2013;11:227.
9 Crowe E, Pandeya N, Brotherton JML, *et al*. Effectiveness of quadrivalent human papillomavirus vaccine for the prevention of cervical abnormalities: case-control study nested within a population based screening programme in Australia. *BMJ* 2014;348:g1458.
10 Baldur-Felskov B, Dehlendorff C, Munk C, *et al*. Early impact of human papillomavirus vaccination on cervical neoplasia-nationwide follow-up of young Danish women. *J Natl Cancer Inst* 2014;106:djt460.
11 Powell SE, Hariri S, Steinau M, *et al*. Impact of human papillomavirus (HPV) vaccination on HPV 16/18-related prevalence in precancerous cervical lesions. *Vaccine* 2012;31:109–13.
12 Walker TY, Elam-Evans LD, Yankey D, *et al*. National, Regional, State, and Selected Local Area Vaccination Coverage Among Adolescents Aged 13-17 Years - United States, 2017. *MMWR Morb Mortal Wkly Rep* 2018;67:909–17.
13 Gertig DM, Brotherton JML, Saville M. Measuring human papillomavirus (HPV) vaccination coverage and the role of the National HPV vaccination program register, Australia. *Sex Health* 2011;8:171–8.
14 Baandrup L, Blomberg M, Dehlendorff C, *et al*. Significant decrease in the incidence of genital warts in young Danish women after implementation of a national human papillomavirus vaccination program. *Sex Transm Dis* 2013;40:130–5.
15 England PH. *Annual HPV vaccine coverage in England: 2012-13*, 2013.
16 Meites E, Szilagyi PG, Chesson HW, *et al*. Human papillomavirus vaccination for adults: updated recommendations of the Advisory Committee on immunization practices. *MMWR Morb Mortal Wkly Rep* 2019;68:698–702.
17 Meites E, Kempe A, Markowitz LE. Use of a 2-Dose Schedule for Human Papillomavirus Vaccination - Updated Recommendations of the Advisory Committee on Immunization Practices. *MMWR Morb Mortal Wkly Rep* 2016;65:1405–8.
18 Krajden M, Cook D, Yu A, *et al*. Human papillomavirus 16 (HPV 16) and HPV 18 antibody responses measured by pseudovirus neutralization and competitive Luminex assays in a two- versus three-dose HPV vaccine trial. *Clin Vaccine Immunol* 2011;18:418–23.
19 Kreimer AR, Rodriguez AC, Hildesheim A, *et al*. Proof-of-principle evaluation of the efficacy of fewer than three doses of a bivalent HPV16/18 vaccine. *J Natl Cancer Inst* 2011;103:1444–51.
20 Romanowski B, Schwarz TF, Ferguson LM, *et al*. Immune response to the HPV-16/18 AS04-adjuvanted vaccine administered as a 2-dose or 3-dose schedule up to 4 years after vaccination: results from a randomized study. *Hum Vaccin Immunother* 2014;10:1155–65.
21 Dobson SRM, McNeil S, Dionne M, *et al*. Immunogenicity of 2 doses of HPV vaccine in younger adolescents vs 3 doses in young women: a randomized clinical trial. *JAMA* 2013;309:1793–802.
22 Brotherton J. Two or three doses of human papillomavirus vaccine? *BMJ* 2015;350:g7778.
23 Niccolai LM, Meek JI, Brackney M, *et al*. Declines in Human Papillomavirus (HPV)-Associated High-Grade Cervical Lesions After Introduction of HPV Vaccines in Connecticut, United States, 2008-2015. *Clin Infect Dis* 2017;65:884–9.
24 Gargano JW, Park IU, Griffin MR, *et al*. Trends in high-grade cervical lesions and cervical cancer screening in 5 states, 2008-2015. *Clin Infect Dis* 2019;68:1282–91.
25 Orenstein WA, Bernier RH, Dondero TJ, *et al*. Field evaluation of vaccine efficacy. *Bull World Health Organ* 1985;63:1055–68.
26 Moulton LH, Wolff MC, Brenneman G, *et al*. Case-cohort analysis of case-coverage studies of vaccine effectiveness. *Am J Epidemiol* 1995;142:1000–6.
27 Clemens J, Brenner R, Rao M. Evaluating new vaccines for developing countries. *JAMA* 1996;275:390–7.
28 Horowitz B, Minor P, Morgenthaler JJ, *et al*. WHO expert Committee on biological standardization. *World Health Organ Tech Rep Ser* 2004;924:1–232.
29 Smith JS, Lindsay L, Hoots B, *et al*. Human papillomavirus type distribution in invasive cervical cancer and high-grade cervical lesions: a meta-analysis update. *Int J Cancer* 2007;121:621–32.
30 Kann H, Hortlund M, Eklund C, *et al*. Human papillomavirus types in cervical dysplasia among young HPV-vaccinated women: population-based nested case-control study. *Int J Cancer* 2020;146:2539–46.
31 National Center for Health Statistics CfDCaP. *Human papillomavirus vaccine impact monitoring project (HPV-IMPACT) overview*, 2018.
32 Xu L, Oštrbenk A, Poljak M, *et al*. Assessment of the Roche linear array HPV genotyping test within the VALGENT framework. *J Clin Virol* 2018;98:37–42.
33 Downs LS, Scarinci I, Einstein MH, *et al*. Overcoming the barriers to HPV vaccination in high-risk populations in the US. *Gynecol Oncol* 2010;117:486–90.
34 Dunne EF, Unger ER, Sternberg M, *et al*. Prevalence of HPV infection among females in the United States. *JAMA* 2007;297:813–9.
35 Hariri S, Unger ER, Sternberg M, *et al*. Prevalence of genital human papillomavirus among females in the United States, the National health and nutrition examination survey, 2003-2006. *J Infect Dis* 2011;204:566–73.
36 Manhart LE, Holmes KK, Koutsky LA, *et al*. Human papillomavirus infection among sexually active young women in the United States: implications for developing a vaccination strategy. *Sex Transm Dis* 2006;33:502–8.
37 Markowitz LE, Hariri S, Lin C, *et al*. Reduction in human papillomavirus (HPV) prevalence among young women following HPV vaccine introduction in the United States, National health and nutrition examination surveys, 2003-2010. *J Infect Dis* 2013;208:385–93.

38 Schiffman M, Castle PE, Jeronimo J, *et al*. Human papillomavirus and cervical cancer. *Lancet* 2007;370:890–907.
39 Stokley S, Jeyarajah J, Yankey D, *et al*. Human papillomavirus vaccination coverage among adolescents, 2007-2013, and postlicensure vaccine safety monitoring, 2006-2014-United States. *MMWR Morb Mortal Wkly Rep* 2014;63:620–4.
40 Wheeler CM. Natural history of human papillomavirus infections, cytologic and histologic abnormalities, and cancer. *Obstet Gynecol Clin North Am* 2008;35:519–36.
41 Oliveira CR, Avni-Singer L, Badaro G, *et al*. Feasibility and accuracy of a computer-assisted Self-Interviewing instrument to ascertain prior immunization with human papillomavirus vaccine by self-report: cross-sectional analysis. *JMIR Med Inform* 2020;8:e16487.
42 Benowitz I, Esposito DB, Gracey KD, *et al*. Influenza vaccine given to pregnant women reduces hospitalization due to influenza in their infants. *Clin Infect Dis* 2010;51:1355–61.
43 Desai SN, Esposito DB, Shapiro ED, *et al*. Effectiveness of rotavirus vaccine in preventing hospitalization due to rotavirus gastroenteritis in young children in Connecticut, USA. *Vaccine* 2010;28:7501–6.
44 Markowitz LE, Hariri S, Unger ER, *et al*. Post-licensure monitoring of HPV vaccine in the United States. *Vaccine* 2010;28:4731–7.
45 Dunne EF, Stokley S, Chen W, *et al*. Human papillomavirus vaccination of females in a large health claims database in the United States, 2006-2012. *J Adolesc Health* 2015;56:408–13.
46 Fleiss JL, Levin B, CP M. *Statistical methods for rates and proportions*. 3rd ed. Hoboken, New Jersey: John Wiley & Sons, Inc, 2003.
47 Lei J, Ploner A, Elfström KM, *et al*. HPV vaccination and the risk of invasive cervical cancer. *N Engl J Med* 2020;383:1340–8.