INSTITUTE OF MEDICINE
OF THE NATIONAL ACADEMIES

Board on Population Health and Public Health Practice

The Committee on Assessment of Studies of Health Outcomes Related to the Recommended Childhood Immunization Schedule commissioned a paper from a consultant, Martin Kulldorff, Ph.D., titled, *Study Designs for the Safety Evaluation of Different Childhood Immunization Schedules*. This paper became available on the project's website on May 14, 2012. The committee invited comments and reactions to the paper to inform the committee discussion. The comment period was open until May 31, 2012.

For more information, please visit the Committee on Assessment of Studies of Health Outcomes Related to the Recommended Childhood Immunization Schedule's webpage: http://www.iom.edu/Activities/PublicHealth/ChildhoodImmunization.aspx.

Please note that any comments that you submit to the committee, including your name and identifying information, will not be kept confidential and will be included in a Public Access File in compliance with Section 15 of the Federal Advisory Committee Act (FACA). The National Academies shall be authorized to use any such comments or submissions in accordance with the National Academies' Terms of Use Statement.

The responsibility for the content of the paper rests with the author and does not necessarily represent the views of the Institute of Medicine or its committees and convening bodies.

Study Designs for the Safety Evaluation of Different Childhood Immunization Schedules

Prepared for the Institute of Medicine Committee on
Assessment of Studies of Health Outcomes
Related to the Recommended Childhood Immunization Schedule

Martin Kulldorff, PhD
Professor, Biostatistician
Department of Population Medicine
Harvard Medical School and Harvard Pilgrim Health Care Institute

May 14, 2012

SUMMARY

To date, there have been few comparative studies evaluating the safety of different vaccine schedules. A few of the existing studies have shown that there are cases in which the risk of adverse events can depend on the vaccination schedule used. Hence, it is both a feasible and important area of study. As a relatively new field of investigation, the big question is what types of study designs will be most fruitful for evaluating different childhood vaccine schedule. A number of possible study designs are presented in this review to evaluate different features or components of the vaccine schedule. These include the timing of individual vaccines; the timing between doses of the same vaccine; the interaction effect between vaccines and concurrent health conditions or pharmaceutical medications; the interaction effects of different vaccines given on the same day; the ordering of different vaccines; and the effect of cumulative summary metrics such as the total number of vaccines or the total amount of some vaccine ingredient. Study designs for the comparative evaluation of one or more complete schedules are also considered. Methods are presented both for adverse events with an early onset, which are the easiest to

study, and for adverse events with a late onset, including serious chronic conditions. It is recommended that a wide variety of different vaccine schedule components should be studied.

1.   INTRODUCTION

Before approval by the Food and Drug Administration (FDA), vaccines are evaluated for efficacy and safety using large phase III randomized controlled trials. For childhood vaccines, the number of children enrolled in these trials is typically in the thousands. That is sufficient to detect common but not rare adverse events. For the latter, there exist several post-marketing vaccine safety surveillance systems using observational data on children who receive the vaccines as part of their general care. In the United States, these include the Vaccine Adverse Event Reporting System (VAERS), the Vaccine Safety Datalink (VSD), and the Clinical Immunization Safety Assessment Network (CISA), all sponsored by the Centers for Disease Control and Prevention (CDC), as well as the Post-Licensure Rapid Immunization Safety Monitoring System (PRISM), which is part of the FDA-sponsored Mini-Sentinel Initiative. Internationally, there are other important vaccine safety surveillance systems such as the Epidemiology Vaccine Research Program at the National Institute for Health Data and Disease Control in Denmark; the Vaccine Adverse Event Surveillance and Communication Network (VAESCO) coordinated by the European Center for Disease Control; the WHO International Drug Monitoring Programme at the Uppsala Monitoring Centre; and the Immunization Division at the Communicable Disease Surveillance Centre in England. All these vaccine safety systems have proved to be very useful and important. They have detected unsuspected adverse events leading to revisions in vaccine recommendations, and in other cases, established the safety of vaccines for which important safety concerns existed. Throughout their existence, there has been continuous and rapid development with respect to the types of questions studied and the epidemiological and statistical methods used. For example, for every new childhood vaccine approved by the FDA, the Vaccine Safety Datalink now conducts near real-time safety surveillance using weekly data feeds from electronic health records (Lieu et al. 2007, Yih et al. 2011). The credit for these continuously improved vaccine safety surveillance systems goes both to the devoted scientists that are building the systems and using them for many important studies,

and, to the vaccine safety advocacy groups that are the key public voice for improved and expanded vaccine safety surveillance.

Most post-marketing studies evaluate the general question as to whether or not a vaccine causes an adverse event among the majority of children receiving the vaccine. Very few post-marketing studies have evaluated whether the risk of adverse events depends on the scheduling of the vaccines. For example, few post-marketing studies have evaluated whether the risk of adverse events depends on the age at which a vaccine is given; on the relative timing of two different vaccines; or on a combined cumulative effect generated by the timing of dozens of different vaccines. These are all different components of the vaccine schedule, and either could potentially be related to the number and severity of adverse events. When evaluating the safety of different vaccine schedules, it is hence important to study the whole range of issues, from the timing of a single vaccine to summary metrics based on the timing of dozens of vaccines.

This is a paper commissioned by the Institute of Medicine (IOM) Committee on Assessment of Studies of Health Outcomes Related to the Recommended Childhood Immunization Schedule. The paper considers different types of potential questions and concerns about the safety of vaccine schedules and describes different epidemiological study designs and statistical methods that can be used to answer such questions in a scientifically rigorous manner. The core of this paper is a set of proposals for the type of study designs and methods that would be appropriate for the comparative evaluation of vaccine adverse events under different vaccine schedules, written in the context of the many difficulties raised by the speakers at the committee meetings held in February and March of 2012. Note though, that it is neither a synthesis, an evaluation, nor a review of the many excellent presentations made at those meetings. Instead, it should be viewed as complimentary information. Note also that the paper does not say anything about the advantages or disadvantages about specific vaccines or vaccine schedules. Rather, the focus is on potential study designs and methods and their ability, or inability, to answer such questions.

1.1   DEFINITION OF KEY TERMS

Potential adverse event: A health event under evaluation in a vaccine safety study, in order to determine if it is caused by the vaccine(s) or not.

Early onset: An adverse event that manifests itself and can be detected within a few weeks after vaccination. May be either acute or chronic.

Late onset: An adverse event that does not manifest itself and/or cannot be detected until a few months or years after vaccination. May be either acute or chronic.

Component of the vaccine schedule: Some specific feature of the vaccine schedule, such as the age at which one of the vaccines is given or the total amount of immune stimulating content received from all vaccines in the schedule. Not to be confused with different components of a single vaccine.

2. VACCINE SCHEDULES, ADVERSE EVENTS AND DATA SETS

2.1 VACCINE SCHEDULES AND THEIR COMPONENTS

To study the safety of different childhood vaccine schedules is an important but complex task. With dozens of vaccines, many of which have multiple doses, there is an almost infinite number of possible vaccine schedules that can be used. To scientifically evaluate the safety of different vaccine schedules, it is necessary to look at specific components of the schedule. Some such components are:

Timing of Specific Vaccines
   a. The age at which a specific vaccine is given, such as the age at the first dose of the hepatitis B vaccine.
   b. The relative timing of different doses of the same vaccine, such as the number of months between the first and second dose of the PCV7 pneumococcal conjugate vaccine.
   c. The interaction between the timing of a specific vaccine and time-varying health events of health status, such as a vaccination given to a child taking a temporary or seasonal medication.

Relative Timing of Two or More Different Vaccines

d. The interaction between different vaccines given on the same day, such as the effect of giving the measles-mumps-rubella (MMR) and varicella vaccines at the same or different health care visit.

e. The order in which different vaccines are given, such as whether measles vaccine is given a few months before or after the diphtheria-tetanus-pertussis (DTP) vaccine.

Cumulative Summary Metrics of a Vaccine Schedule

f. The total number of vaccinations given to the child before a certain age, such as the 6th birthday.

g. The average age at which the vaccines were given.

h. The cumulative amount of immune stimulating content present in all vaccines received.

In addition to specific components of the vaccine schedule, one can also try to compare complete vaccine schedules.

Comparison of Complete Vaccine Schedules

i. Whether the child has approximately followed the CDC recommended vaccine schedule, or not.

j. The comparative safety of a specific alternative vaccine schedule, such as Dr. Bob's (Sears 2007), versus the one recommended by CDC.

The study design and statistical methods used depend on which vaccine schedule component that is being evaluated. As they are quite different, each component (a) to (e) is dealt with in separate sections of this paper, from section 3 to 7. For cumulative summary metrics, the methods are similar irrespective of what component of the vaccine schedule the metric is designed to measure. Hence, components (f) to (h) are treated together in section 8. In section 9, methods for comparing different complete vaccine schedules are discussed. More general methodological issues are discussed in section 10.

The different types of studies should not be done in isolation from each other. If it is found that one complete vaccine schedule has an excess number of adverse events compared to another, we do not know which component of the schedule caused the

difference. Hence, it is not recommended to conduct studies comparing complete schedules without also evaluating specific components of those schedules. Likewise, when studying a specific component, results may be confounded by other components of the vaccine schedule. For example, a child receiving vaccine A at an early age may be more likely to also receive vaccine B at an early age, and the timing of vaccine A will then be correlated with the number of adverse events even if it is the timing of vaccine B that is the culprit. It could also be that there are two different vaccine schedule components that cause adverse events, but that they cancel each other out when looking at the difference between two complete schedules, making it impossible to detect the problem if only the complete schedules are studied.

Another reason for studying specific components of the vaccine schedule is that, if a problem is found, we need to know how to revise the schedule in order to reduce the number of adverse events. Just because one complete vaccine schedule is found to cause more adverse events than another, we do not necessarily have to revise all components of that schedule.

2.2   ADVERSE EVENTS WITH EARLY VERSUS LATE ONSET

In vaccine safety studies, the goal is to evaluate if there is a causal relationship between the vaccine(s) and some health event of interest. The latter is denoted as a potential adverse event, as it may or may not be an actual adverse event caused by the vaccine(s). The type of health event under study determines the appropriate methodological methods for vaccine safety studies. This paper considers two main types. The first type consists of potential adverse event with an early onset, and which can be detected soon after the onset. The event itself could either be acute, of a passing nature without any permanent damage, such as a febrile seizure; or chronic, lasting many years, such as a stroke. The second type consists of potential adverse events with a late onset, several months or years after vaccination, and events with an early or a gradual onset that cannot be detected until long after vaccination. For simplicity, all of these are denoted as 'late onset'. These potential adverse events can also be either acute or chronic in nature.

The most suitable study design and analysis methods are greatly dependent on whether the potential adverse event has an early or late onset, and in the

description below, separate methods are proposed for the two outcome types. This is a little bit of a simplification, since there are of course also potential adverse events that fall somewhere in between on this spectrum. It should also be pointed out that an early onset chronic condition can be studied using either the methods described for early or late onset, but the early onset methods are in most cases preferable.

Another key issue is whether there is a clear time at which the potential adverse event happened, as with, for example, a seizure; or whether the disease evolves more gradually, without a single clearly defined day of onset, as with, for example, narcolepsy or autism. This does not affect the study design as much as the time of onset, but it is an important consideration when defining and collecting the data.

For most potential adverse events, we are only interested in incident diagnoses. That is, the first time that particular diagnosis has been made. For example, if a child is diagnosed with asthma at age 2, and then has a follow-up visit for his/her asthma at age 4, we do not want to attribute the asthma to a vaccination given at age 3. Depending on the potential adverse event under study, one can define an incident diagnosis as a diagnosis that has not occurred during the previous D days. The value of D will depend on the adverse event, but a typical value is about one year.

The potential adverse event studied can either be very specific, such as febrile seizures or autism, or it could be more general, such as all cause outpatient physician visits, emergency department visits, or hospitalizations. The latter may seem more desirable, as it includes the combined effect of the vaccine schedule on all important health events, but the opposite is true. Such general definitions are more prone to biases, and they are therefore more difficult to study. This is because people that follow the CDC recommended vaccine schedule may be different from those that do not, in terms of their health care seeking behavior, and especially those that do not deliberately postpone their child's vaccinations. For example, parents that are more prone to take their children to the doctor when the child is sick may also be more prone to take their children to the well care visits during which most vaccines are given.

2.3   DATA SETS FOR POST-MARKET VACCINE SAFETY STUDIES

To facilitate the understanding of the study designs and methods described in subsequent chapter, a brief background is first given concerning some of the data sets that are available and currently used for post-marketing vaccine safety studies.

Pre-Market Randomized Trials: Phase III randomized trials are primarily designed to evaluate the efficacy of vaccines. They are also able to find common adverse events, but their sample size is typically not large enough to evaluate rare but serious adverse events. Their primary use for post-market vaccine safety surveillance is to generate study hypotheses. For example, a single case of Kawasaki disease in the vaccine arm of a phase III randomized trial is not evidence that the vaccine causes Kawasaki disease, since it could be pure coincidence, but it may warrant a post-marketing safety evaluation.

Spontaneous Reporting Systems: Most countries in the world have a vaccine safety surveillance system based on spontaneous reports. These are linked together through the WHO Collaborating Centre for International Drug Monitoring in Uppsala, Sweden, so that it is possible to combine data from multiple countries. In the United States, CDC and FDA are joint sponsors of the Vaccine Adverse Event Reporting System (VAERS).

These systems contain spontaneous reports of suspected vaccine adverse events, sent in by physicians, nurses, patients, parents, manufacturers, and others. The gender and age of the vaccinated person are some of the variables collected. There is often information about multiple vaccines given on the same day. Analyses are done using proportional reporting ratios (Evans et al. 2001) and similar methods. For example, if 1.5% of all vaccine related adverse event reports are for seizures, and there are 1,000 reports for vaccine A, then we would expect 15 seizure reports for vaccine A. If, in reality, there are 45 such reports, the proportional reporting ratio is 3. That is more than what one would expect, and it may indicate that there is an excess risk of seizures after vaccination. Actual analyses are more complex, since it is necessary to adjust for age and other variables. There are also other more sophisticated methods used (Bate et al. 1998, DuMouchel 1999, Rothman et al. 2011).

The major advantage of VAERS is that it receives reports from the whole country. The two major disadvantages are that there is underreporting and that there are no reliable denominator data. That is, while we have information about a number of vaccinated children with the potential adverse event of interest, we do not know the total number of children that were vaccinated, how many unvaccinated children had the same type of event, or how many vaccinated children had the event without it being reported.

Some reports to VAERS are studied further in the Clinical Immunization Safety Assessment Network. Among other things, this network aims to '*improve the scientific understanding of vaccine safety at the individual patient level*', by obtaining and evaluating detailed genetic and other information from each patient (LaRussa et al. 2011).

Electronic Medical Records: For 2011, it is estimated that 57% of office based physicians used electronic medical records (EMRs), up from 24% in 2005 (Hsiao et al. 2011). The EMRs most useful for medical research are the ones from large health plans as they contain medical records for a well-defined member population, including both inpatient and outpatient encounters. The Vaccine Safety Datalink (VSD) project is the premier EMR-based vaccine safety system in the United States (Chen et al. 1997, DeStefano 2001, Baggs et al. 2011). Led by CDC, it is a collaboration with ten health plans: Group Health in the State of Washington, Harvard Pilgrim / Atrius Health in Massachusetts, HealthPartners in Minnesota, Kaiser Permanente in Colorado, Georgia, Hawaii, Northern California, Oregon and Southern California; and Marshfield Clinic in Wisconsin. Together, these health plans have around 9.5 million members and an annual birth cohort of over 100,000. The VSD system is used both for retrospective studies and for near-real time vaccine safety surveillance with weekly analyses of newly approved vaccines. Similar systems exist in a few other countries, including the Epidemiology Vaccine Research Program at the National Institute for Health Data and Disease Control (Seruminstitutet) in Denmark

The major advantage with EMR systems is that denominator data is available, as all vaccinated children can be identified. It is then possible to compare the number of adverse events in vaccinated and unvaccinated children, or, vaccine exposed and

unexposed time periods within the same child. A disadvantage is that data is registered for purposes other than research, and there is sometimes miscoding of health events. Depending on the health outcome, manual chart review is therefore sometimes warranted.

Health Insurance Claims Data: Health insurance companies have medical information for millions of insured members and their families, which they receive when doctors and hospitals file their financial reimbursement claims. One such system in the United States is the Post-Licensure Rapid Immunization Safety Monitoring (PRISM) program, run by the Food and Drug Administration as part of their Mini-Sentinel project (Nguyen et al. 2012). Claims data is more limited than electronic health records, but can be used in much the same way for post-market vaccine safety studies. The major advantage is the large sample size that can be achieved. The major disadvantage is that some health conditions are not captured. Depending on the potential adverse event under study and the confounders that need to be adjusted for, this may or may not be a problem.

Because of their similarities, electronic medical records and health insurance claims data will; be treated as the same type of data in this report, under the name of *health plan data*.

Study Specific Data Collection: Sometimes new data is collected specifically for vaccine safety studies, such as a self-controlled case series, a case-control study, a cohort study, or a post-market randomized trial. An intermediate option is to obtain some of the data from health plans, disease registries and/or vaccine registries, while the remaining data is collected from study specific patient surveys or measurements. The available options are too many to provide a detailed description of each.

3.   TIMING OF SPECIFIC VACCINES

In a randomized childhood vaccine trial, the age at which the vaccine is given is tightly controlled by the study design, to correspond to the future planned vaccine schedule. This is appropriate, but once a vaccine is on the market, it is also given at a

wide variety of other ages, for a variety of reasons. There are two scenarios in which it is of great interest to evaluate the risk of a vaccine as a function of the age at which the vaccine was given. (i) If a vaccine safety study has shown that there is a statistically significant excess risk of an adverse event, we want to know if the excess risk varies by the age at which the vaccine was given. (ii) Even if a general safety study covering all age groups has not shown a statistically significant excess risk of the adverse event, there could still be an excess risk if the vaccine is given at certain ages outside the recommended schedule. Such a safety problem could be masked by the non-effect among the most populous age group, and a special study looking at age-specific risks would be warranted.

3.1   KNOWN ADVERSE EVENTS WITH EARLY ONSET

Background: Some vaccines have been shown to cause an acute adverse event to occur within a few weeks after vaccination. Examples include intussusception 3-7 days after RotaShield vaccination (Kramarz et al., 2001; Murphy et al. 2001) and febrile seizure 7-10 days after MMR and measles-mumps-rubella-varicella (MMRV) vaccination (Klein et al. 2010). There are also several such examples of less severe adverse events like fever and rash. The adverse event may be serious enough to warrant the withdrawal of the vaccine from the market, as with RotaShield, or it may be mild enough to keep using the vaccine, as with MMR. A mid-level alternative option is to revise the vaccination schedule to minimize the number of adverse events, or, to contraindicate the vaccine in certain age group. Knowledge of the relative and attributable risk of the adverse event as a function of age is one important component when deciding between these options, together with other important factors such as how the immunogenicity varies by age. This paper only discusses methods for obtaining knowledge about the former, and not how to weight different sources of information to arrive at a final decision.

Examples: In two different studies, Murphy et al. (2006) and Rothman et al. (2006) evaluated the effect of age on the excess risk of intussusceptions after RotaShield vaccine. In a more recent study, Rowhani-Rahbar et al. (2012) evaluated the effect on age on the risk of febrile seizures after MMR and MMRV vaccinations. All three studies found that the risk of the adverse event varied greatly by age.

Data: Electronic medical records from health plans or health insurance claims data are ideally suited for studying this question. It is also possible to use data from a case-control study. VAERS data cannot easily be used since it does not contain information about the age distribution of vaccinated children. Data from pre-marketing randomized trials are too small and typically do not include a wide enough range of ages. In light of existing observational data, specifically designed post-marketing randomized trials could be unethical depending on the nature of the known adverse event.

Method: The first key step is to determine the time between the vaccination and the adverse event as precisely as possible. Some children will, just by chance, have the adverse event soon after vaccination. To maximize the precision of our age estimates, we want to exclude as many of them as possible, by only counting the adverse events occurring in the true risk window. An efficient way to determine the appropriate risk window is to use a temporal scan statistic. For a cohort of vaccines with a subsequent event of interest, record the number of days from vaccination to the event. Ignore events that occur on the same day as the vaccination, as they may have a different background rate, as well as those that occur beyond an upper limit such as 70 days after vaccination. If there is no relationship between the vaccine and the adverse event, we expect the adverse events to be uniformly distributed during the [1,70] day period. The temporal scan statistic scans the time period for any cluster of events, without any assumptions about their location or length. The method determines the statistical significance of such clusters adjusting for the multiple testing inherent in the hundreds of overlapping time periods evaluated. As an example, temporal scan statistics were used to determine that the excess risk of seizures after MMRV vaccination is confined to the 7-10 day post-vaccination period (Klein et al. 2010).

The second step is to evaluate the relationship between age at vaccination and excess risk of the adverse event. The simplest and most common way to do this is to divide age into different age groups, such as 6-12 months and 12-24 months, and compare the risk. Since it is unrealistic to assume that the risk suddenly jumps at a particular age, and for greater precision, it is better to model risk as a continuous function of age. This can be done using either regression with first, second and higher degree polynomials or by using regression splines (Rothman et al. 2006).

In these analyses, it is important to take the underlying natural age-related risk into account. For example, the incidence of intussusceptions is very low immediately after birth, after which it gradually increases until about 5 months of age, after which it gradually decreases (Eng et al., 2012). There are a number of possible ways to adjust for this, depending on the exact study design. In a cohort study of vaccinated individuals, one can use historical data to estimate the age curve, using a polynomial function, and then use that as an offset term in the regression model. An alternative approach is to use both a risk and control interval for each individual, in a self-controlled analysis, evaluating whether the difference relative risk in the risk and control windows are different by age. Note though, that if the natural incidence rate for the adverse event vary greatly by age in weeks rather than year, it is still necessary to incorporate an offset term based on the natural age curve even when a self-controlled analysis is conducted. In a case-control study, matching by age ensures that the age-based incidence curve is adjusted for.

3.2   VACCINE RISK FOR SPECIFIC AGE GROUPS: EARLY ONSET AVERE EVENTS

Background: Most childhood vaccines are given according to the recommended schedule, but some children may get the vaccine at a much earlier or much later age. There are many potential reasons for this, including high risk of exposure due to a current disease outbreak or due to family members with the disease, missed well care visits, shortages of the vaccine, parental or physician concerns about the recommended vaccine schedule, misunderstanding of the recommended schedule, or medical errors. For example, while the first dose of MMR is recommended at age 12-15 months, in one health plan 22% of children receive it later and 0.7% percent receive it before their first birthday, with 0.3% before 6 months of age. Nationwide, even half a percent adds up to a fairly large number, and it is important to evaluate the safety of the vaccine for those children, so that a contraindication warning can be issued if there is a major safety problem.

Example: After the 2004 recommendation to give influenza vaccines to 6-24 month old children, Hambridge et al. (2006) used data from the Vaccine Safety Datalink to

conduct an influenza vaccine safety study specific to this age group, looking at a wide variety of potential adverse events.

Data: Electronic health plan data captures all vaccinations at whatever age they occurred, so such data is not only useful for evaluating the safety of vaccines in special age groups, but also to characterize the real world age distribution of vaccinated children.

With its national coverage, VAERS data can also be used to monitor the vaccine safety in specific age groups. While no denominator data is available directly, a large number of adverse event reports in children outside the recommended age range could be the first indication that an age-specific problem exists.

If a change is anticipated in the recommended age that a vaccine should be given, a randomized trial may be warranted. For that to occur, there needs to be some uncertainty as to whether the currently recommended time is safe, and some evidence that an alternative age is safer, based on, for example, observational data. If the question is simply whether the vaccine should be contraindicated for certain age groups, or whether the vaccine is also safe outside the recommended age of vaccination, without any evidence of harm at the time recommended age, then a randomized trial would not be ethical.

Methods: For health plan data there are a few different analysis options. For early onset events, a self-controlled risk interval design can be used. First, decide on a risk window, such as 1 to 21 days after vaccination, and a control window, such as 22 to 42 days after vaccination. For each vaccinated child in the age group of interest, count how many of them had an adverse event in the risk and control windows respectively. Suppose that the two windows are of the same length, and that there are a total of *n* adverse events in the two windows combined. The number of adverse events in the risk window then has a binomial distribution with parameters *n* and ½.

Since this is a self-controlled analysis, it is only time-varying confounders that may need to be adjusted for, and there is no need to worry about gender, genetics, stable environmental factors, study site, etc. For some adverse events where the

14

incidence rate changes rapidly from one week of age to the next, an age adjustment must be made. If the age distribution of the disease is know, this can easily be done by using an offset term in a logistic regression model. The same is true if there are strong seasonal trends in the incidence rate. An alternative way to adjust for seasonality is to use a case-centered approach, as proposed by Fireman et al. (2009).

The choice of risk and control windows depends on the vaccine and the adverse event. Sometimes it is worthwhile to have a wash-out period between the two windows. To avoid day-of-week effects, the two windows should have the same number of days in any modulus of seven. For example, the risk window may be 1 to 14 days and the control window 22 to 70 days; or the risk window may be 1 to 2 days and the control window days 8 to 9 together with days 15 to 16. Theoretically, it is also possible to use a comparison window before vaccination, but that can introduce confounding by indication or contraindication.

In VAERS data, the age of the vaccinated child is one of the variables collected. To evaluate whether a vaccine is safe outside the recommended schedule, it is hence possible to look at specific pre-defined age groups. This can be done using the same methods that are used for all age groups combined, such as proportional reporting ratios (Evans et al. 2001).

4.   TIME BETWEEN VACCINE DOSES

Background: Almost all childhood vaccines are given in multiple doses a few months or years apart. It is conceivable that the length of the time interval between vaccine doses could increase or decrease the risk of adverse events.

Example: Using a randomized trial, Pitman (2002) showed that the risk of adverse events was reduced if the second dose of subcutaneous anthrax vaccine, adsorbed, is given 4 rather than 2 months after the first dose.

4.1   EARLY ONSET ADVERSE EVENTS

Data: Electronic health data are suitable to use for vaccine doses that are at most a few years apart. If the time between doses is too long, health plan data is less suitable as only some members will have been enrolled long enough to have information about all the doses of interest.

Method: For simplicity's sake, first consider the situation where we want to evaluate the length of the time interval between the first two doses of the vaccine with respect to early onset adverse events after the second dose. First, identify a cohort of children who received the first two doses of the vaccine. Exclude children that do not have a sufficiently long enrollment in the health plan to ensure that these are truly the first two doses. Note the number of days between the doses, and whether they had an adverse event during a pre-specified risk window after the second dose. For the statistical analysis, use logistic regression. The dependent variable is whether the potential adverse event was present in the risk window or not. The independent variable of interest is the number of days between the two doses. Adjust for gender, age at the second dose, calendar year, seasonality, study site, and any other potential confounders by including these as additional independent variables.

When looking at early onset adverse events after the third dose, the same methods can be used for evaluating the length between the first and third dose or between the second and third dose. A single logistic regression can be used to evaluate both the time from the first to the third dose and the time from the second to the third dose, by including both of them as two separate independent variables. The same applies for early onset adverse events after subsequent doses.

If an excess risk is found, it is not clear from this design whether it is an excess risk due to the time length between the vaccination, or, if there is an excess risk driven purely by the first dose that showcases itself after the second dose in children that receive it sooner. By estimating the temporal function of any excess risk due to dose one alone, this can be adjusted for, either as an offset term or by including it as an additional variable in the regression model, which then also has to include unvaccinated children in the same age group. An alternative, simpler approach, is to limit the study to children where the doses are given at least X days apart, where X is chosen to be large enough that it is unlikely that there is any excess risk beyond that time that is purely due to the first vaccine.

If there is some evidence from the observational study that there is a differential risk depending on the time between vaccinations, but it is not conclusive, then a randomized trial could be conducted. For example, in a study of a rotavirus vaccine, children may be randomized to receive the three doses at age 2, 4 and 6 months of age, according to the CDC recommended vaccine schedule, plus a placebo dose at age 9 months; versus three doses at age 2, 6 and 9 months, with a placebo dose at age 4 months. The results from the observational study, with its wide variety of schedules, can be used to inform the definition of the study arms in the randomized trial. Note though, that if the adverse event is rare, a randomized trial is not a feasible approach, as the required sample size would be very large, and hence, the cost of the trial prohibitively expensive.

4.2   LATE ONSET ADVERSE EVENTS

Data: Electronic health data are suitable to use for late onset adverse events that occur within a few year after the last dose, and for vaccines for which all doses of interest are given within a year or two. For late onset events and longer times between doses, health plan data may be less suitable as only some members will have been enrolled long enough to be informative.

Method: For late onset adverse events, the same methods can be used as for early onset events, with some modifications. Most importantly, rather than defining adverse events in terms of a pre-defined risk window after the last dose, it is more suitable to define them according to a pre-defined age range. For example, the study may only include children who had all the doses of interest before 18 months of age, and would only consider adverse events for which the incident diagnosis occurred between ages 2 and 6.

If suitable health plan data is not available, a case-control approach can be used instead. The first step is then to select children with an incident diagnosis during a pre-defined age range, together with a set of controls matched by age, gender, calendar month, study site and other covariates of interest. The next more challenging step is to obtain the vaccination history of each of the children. This could be done by contacting all the health plans or all the pediatricians that the child

has had. The dose interval length is then compared between those children with and without the adverse event.

5.   INTERACTION EFFECTS BETWEEN VACCINES AND HEALTH CONDITIONS

Background: Several vaccines are contraindicated for children with specific health problems. For example, live attenuated influenza vaccine should not be given to 2 to 4 year old children who have had wheezing during the past 12 months. This means that the vaccine schedule may have to be modified for some children based on their personal disease history. To know if that is necessary, one needs to study the interaction effects between vaccines and pre-existing health conditions.

The study of interaction effects between vaccines and health conditions is especially important when there is a known excess risk of an adverse event. If it is possible to pinpoint that the adverse events are due to an interaction effect, then the number of adverse events can be reduce by contraindicating the vaccine to children with the health condition in question. For example, if the risk of seizures after MMR is higher among children with a recent well-defined disease episode, then the MMR vaccine may potentially be postponed by 3 months for those children.

5.1   EARLY ONSET ADVERSE EVENTS

Data: Electronic health plan data are suitable for this question.

Method: First consider the scenario in which there is a known increased risk of the adverse event in the population as a whole, and we want to know if the excess risk is more severe among a specific group of children, Using the temporal scan statistic, first determine the true risk window for the adverse event, as described in Section 3.1 above. Suppose we have an excess risk in the 7-10 days after vaccination. We now define the study population as those who received the vaccine, and, who had an adverse event in some longer time period, such as 1 to 42 days after vaccination. In a logistic regression model, the dependent variable is whether they had the adverse event inside or outside the true risk window. The independent variables are the various health status variables that we want to examine as potential risk

18

modifiers. Several of these can be included in the same logistic regression, but doing several univariate analyses may be a suitable first step. As long as the baseline risk for the adverse event is fairly constant over the longer time period, it is not necessary to adjust for age. Note that, since all subjects had the vaccine and all subjects had the adverse event, there is no actual interaction term in the logistic regression model.

If we do not have a known adverse event, but still want to evaluate possible vaccine-health status interaction terms, we can still use the same approach with a reasonable guess of a wider risk interval. For example, the risk interval may be 1 to 42 days while the comparison interval is 42 to 84 days.

5.2   LATE ONSET ADVERSE EVENTS

The approach described above cannot be used to study late onset adverse events. Instead, we first determine a vaccinated child also had the health condition of interest at the time of vaccination. We then compare the number of late onset events between the children that did and did not. This design is more prone to bias than the self-controlled design described above. One way to partially adjust for this is to only include children that both had the vaccine and the potential adverse event of interest, at any time, and compare the children who had the vaccination at the same time as the health event with those that had them at different times.

6.   VACCINE-VACCINE INTERACTION

Background: In the CDC recommended vaccine schedule, many different vaccines are given on the same day. It is plausible that two vaccines, if given separately from each other, do not increase the risk of adverse events, but if they are given on the same day, there is a vaccine-vaccine interaction effect, leading to increased risk. It could also be that one or both of the vaccines, when given separately, leads to a modest excess risk of the adverse event, bur when given together, leads to a much higher excess risk.

Example: With data from the Vaccine Safety Datalink and separate self-control risk interval analyses, it was found that there was an increased risk of seizures 1 to 2 days after trivalent inactivated influenza vaccine (TIV), and also, that there was an increased risk of seizures 1 to 2 days after the PCV13 pneumococcal conjugate vaccine. To tease apart the effects from the two different vaccines, and to evaluate the interaction between the two, the author of this paper suggested the approach mentioned below and worked out the formulas for the analysis. It was found that both vaccines had caused an excess risk of seizures 1 to 2 days after vaccination, irrespectively of the presence of the other vaccines, and that the effects were independent of each other (Tse et al., 2012). This means that there was a positive additive interaction but no multiplicative interaction. Hence, the estimates obtain indicated that it is safer to give the two vaccines on separate days rather than on the same day.

6.1   EARLY ONSET ADVERSE EVENTS

Data: Both VAERS and electronic health plan data can be used to evaluate early onset adverse events due to vaccine-vaccine interaction.

Method: For spontaneous reporting systems, Almenoff et al. (2003) have developed a proportionality-based version of DuMouchel's (1999) empirical Bayes multi-item gamma Poisson shrinker. Pairs of two drugs are treated as a separate unique drug, different from individual drug users. To signal as a possible interaction-induced adverse event, the lower 5% percentile of the empirical base geometric mean estimate must be larger than the upper 95th percentile of the empirical base geometric mean estimate for both of the individual drugs. This approach makes sense in a data mining context, where it is necessary to have some form of formal or informal adjustment for the multiple testing.

Two other methods for evaluating interaction effects in spontaneous reports have been proposed by Thakrar et al. (2007) and Norén et al. (2008). Both of these, as well as the previously described method by Almenoff et al. (2003) were proposed for drug-drug interactions, but they can also be used for vaccines.

For electronic health plan data, a different methodological approach is needed. With a self-controlled risk interval analysis, it is possible to evaluate the effect of a vaccine on an adverse event by comparing the number of adverse events in a risk interval right after the vaccine is given versus a control interval long after vaccination. Now, suppose we have two vaccines, such as PCV13 and TIV, and we want know to if there is an increased risk of seizure 1 to 2 days after vaccination. We can then do a self control risk interval analysis for TIV and another one for PCV13, ignoring any other vaccines given on the same day. Suppose that we see an excess risk in both analyses. Since these particular vaccines are often given on the same day, it then is not clear if:

i. TIV causes an excess risk, and PCV13 is just an innocent bystander, with no excess risk.
ii. PCV13 causes an excess risk, and TIV is just an innocent bystander, with no excess risk.
iii. Both vaccines cause an increased risk of seizures independently of each other.
iv. There is either a positive or negative interaction effect between the two vaccines.

To better understand the importance of different interaction effects, a few examples are given. Assume that TIV alone causes a two-fold excess risk and that PCV7 alone also cause a two-fold excess risk:

i. If, when taken together, they also cause a two-fold excess risk, then there is negative interaction, and it is safer to take the two vaccines on the same day.
ii. If, when taken together, there is a three-fold excess risk, then there is negative multiplicative interaction, while there is no additive interaction. In this scenario, it is equally safe to take the two vaccines on the same or on separate days.
iii. If, when taken together, there is a four-fold excess risk, then the two vaccines act independently on the multiplicative scale (i.e., TIV doubles the risk, and then PCV7 doubles the risk on top of that), while there is a positive interaction on the additive scale. In this scenario, it is safer to give the two vaccines on separate days, since one time period with a four-fold excess risk is worse than two time periods with two-fold excess risk.

21

iv. If, when taken together, there is a five-fold excess risk, then there is positive interaction on both the multiplicative and the additive scale, and again, it is safer to take the vaccines on separate days.

While it is possible to do three separate self-controlled risk interval analyses for the vaccines when given alone and when given together, the best approach is to combine all the information into one logistic regression model that includes both the main effects and the interaction terms. This can be done, and it is then possible to formally test for a multiplicative interaction effect.

7. VACCINE ORDER

Background: While not a common concern, it has occasionally been suggested that the order in which vaccines are given may be influence the risk of adverse events. Here, we are not thinking of vaccines given a couple of minutes apart at the same health care visit, but of vaccines given a few days, weeks or months apart. For example, in a study of DTP and measles vaccines in a low-income African country, Aaby et al. (2003) hypothesized that '*DTP as the last vaccine received may be associated with slightly increased mortality*'. Veirum et al. (2005) suggested that '*it might be examined whether provision of BCG or measles vaccine shortly after the last dose of DTP could secure specific protection and prevent the negative immune stimulation associated with having received DTP*', and that '*different sequences of vaccinations*' might have to be considered.

Example: In a three-arm randomized vaccine trial, with a total of 1027 children, Leonadri et al. (2011) compared both immunogenicity and safety for different orders of the MMRV and PCV7 vaccines. In the study, the MMRV vaccine was given either 6 weeks prior, on the same day, or 6 weeks after the fourth dose of the PCV7 vaccine. The incidence of local and systemic adverse events was comparable between the groups while no serious adverse events were reported in either group.

Data: Electronic health plan data are ideally suited to study this question. Alternatively, children with the outcome of interest could be identified using, for

example, hospital data or a disease registry, followed by a vaccination history survey to their parents.

7.1   LATE ONSET ADVERSE EVENTS

Method: The following is a study design for comparing the order of vaccines A and B, using health plan data. For simplicity, this description assumes a single dose of each vaccine, but it can be generalized to multiple doses. First, identify children with the adverse event outcome of interest. In the study, include only those children that had both vaccines A and B at least X days prior to the onset of the disease, except those that had both vaccines on the same day. With health plan data, the comparison group will be all other children. For each child, note the timing of the two vaccines. For each child with the adverse event: (i) note the exact age at disease onset, (ii) calculate how many of the comparison children of the same gender that also had both vaccine A and B at least X days before that age, but not on the same day, and (iii) note the proportion of those that had A before B. This proportion is the estimated probability that the study child had vaccine A before B, under the null hypothesis of no effect of vaccine order. The analysis is adjusted for age and gender, and other covariates can be adjusted for in the same way as gender.

The reason for excluding children that had one of the vaccines less than X days before the adverse event is to remove the effect of vaccine specific early onset adverse events that is caused by one of the vaccines independently of the presence of the other. The value of X will depend on the vaccines and adverse event studied. To avoid bias, it should be large enough so that the any adverse events caused by one vaccine independently of the other do not vary in time based on the number of days after vaccination.

An alternative is to use a case-control design. The children with the adverse event are selected as before. Second, identify a comparison group of children who did not have the adverse event outcome, matched by age, gender and any other variables, and with the same inclusion criteria, with respect to vaccination history. Then compare how many of the cases and how many of the controls had vaccine A before vaccine B, and vice versa.

7.2   EARLY ONSET ADVERSE EVENTS

Method: The above design cannot be used for acute adverse events, due to the bias mentioned above. Instead, the following design can be used. Using health plan data, select children that had vaccine A at any time. Separate them by whether they have had vaccine B prior to A, or not. The compare these two groups in term of how many of them had the adverse event of interest on the 1 to D days after vaccine A. The value of D defines the risk window, and will depend on the vaccines and adverse event studied. The analysis can be performed using unconditional logistic regression, adjusting for covariates such as age at vaccination, gender, calendar years, study site, etc.

This study design cannot in itself distinguish between an effect due to the order of the vaccines and an interaction effect, where the risk increases with the same amount after the second vaccine irrespectively of their order. By collecting the above data for both vaccine A and then in the corresponding manner, for vaccine B, it is possible to compare the two risks estimates. If the increased risk is due to vaccine-vaccine interaction, but not the order of the vaccines, these estimates should be the same.

8.   CUMMULTIVE SUMMARY METRICS FOR VACCINE SCHEDULES

Background: It is conceivable that it is neither the timing of individual vaccines nor the interaction between vaccines that are responsible for adverse events, but rather, some more general component of the vaccine schedule, such as the total number of vaccines given; the average age at which the vaccines are given; or the cumulative amount of immune-stimulating content, immunogenic adjuvants or preservatives in all vaccines received. Similar study designs and statistical methods can be used for most of these types of metrics, so they are considered together.

Examples: Total amount of immunogens exposed to, and the maximum amount of immunogens exposed to on a single day, has been used in studies of autism (DeStefano et al., 2012) and neuro-psychological outcomes (Iqbal et al., 2012).

Summary Metrics for General Features of Vaccine Schedules: The first and most critical step is to define one or more suitable metrics reflecting the general feature of the vaccine schedule that should be evaluated. The number of options is large. Here are some examples:

- The total number of vaccines received before the child's 6th birthday.
- Average age at which a specified set of vaccines were given. Children who did not receive all vaccines can either be excluded from the study, or, one can take the average age of the vaccines given, or, ages may be imputed for those vaccines that were not received.
- The age at which a certain portion of the recommended vaccine schedule was completed. For example, one could take the set of vaccines that CDC recommends for the first 18 months, and determine the age at which all of them have been given. Children that did not get all the recommended vaccines could either be excluded from the study, or, an upper limit on the age may be imputed for those children that have not received all vaccines by that age.
- The average number of vaccines given at each visit. For example, if one child had 15 vaccines spread out over 5 visits, the average is 3. If another child had only 3 vaccines in total, all given at the same visit, the average is also 3.
- The total amount of immunogens (antibody-stimulating proteins and polysaccharides) exposed to from all vaccines combined (DeStefano et al. 2012).
- The total amount of immunogenic adjuvants exposed to from all vaccines combined.
- The total amount of the thimerosal preservative exposed to from all vaccines combined (Price at al, 2010).
- The number of days undervaccinated (Glanz et al. 2012). For each vaccine, calculate the number of days between the recommended age and the actual age of vaccination, and then sum over all vaccines. For a perfectly compliant child, the value is zero.
- The maximum number of vaccines received on a single day, or, the maximum amount of immunogens or adjuvants received on a single day.

Neither of these definitions is the 'right one', and they serve as examples of what could be used rather than recommendations of what should be used. The choice will and must depend on the scientific hypothesis that the scientist is evaluating.

All of the above metrics are continuous or ordinal in nature. Each one could be dichotomized into a 0/1 variable. For example, in the first example above, the children could be split into those receiving at least 10 versus less than 10 vaccines. Such dichotomization is not recommended though. If there is a difference in risks between receiving 9 and 10 vaccines, there is probably also a difference between 5 and 9 vaccines and between 10 and 15 vaccines. Such information is thrown away when the data is dichotomized, and hence, statistical power is lost.

A key criterion when deciding between these metrics is whether one wants to compare the recommended vaccine schedule with children who receive the same set of vaccines, but through a different schedule, or, if one wants to compare children that receive all versus only some of the recommended vaccines. These are two fundamentally different questions. The first one is exclusively focused on how the vaccinations are scheduled. With the latter option, any differences seen in the risk of adverse events could either be due to the how the vaccines are scheduled, or more likely, on the different sets of vaccines received, and without further studies, it is impossible to know which.

Data: Electronic health plan data provide one of the best opportunities to study the safety of vaccine schedules with respect to cumulative summary metrics. As the complete vaccination history is needed to calculate the metric of interest (such as average age at vaccinations), population based studies will be limited to children with a sufficiently long enrollment in the same health plan. In some European countries such as Denmark, with a national health care system, these studies are easier to conduct, as only a small percent of children immigrate to or emigrate from the country. From the United States perspective, a drawback of doing these types of studies in foreign countries is of course that their recommended vaccine schedule is different from ours.

Methods: Once the outcome definition has been decided, the methods one can use for these more general vaccine schedule components are very similar to those

described in Section 4.2 concerning the time between vaccine doses. With health plan data, first classify each child according to one or more of the metrics above. Only include children with a sufficiently long enrollment period. As the next step, determine if they have the adverse event of interest during some pre-defined age period. For the statistical analysis, use logistic regression with the potential adverse event as the dependent variable and the vaccine schedule metric as the independent variable. More than one metric for the vaccine schedule can be included in the same regression model, by which it is possible to try to tease apart the relative influence of each one on the adverse event. Gender, calendar year, study site and other covariates can be adjusted for by also including them as independent variables in the logistic regression.

To include as many children as possible in the study, irrespectively of their length of enrollment, a Cox proportional hazards survival analysis model could be used instead of logistic regression. Children leaving the health plan are then censored at that time. Note though, that the enrollment period must still be long enough to calculate determine their vaccine schedule in sufficient detail to calculate the metric of interest.

The key strength of health plan data is the availability of detailed longitudinal vaccination and disease history for millions of children. It does not contain all potential information of interest though. If some of the exposure history is unavailable, such as the particular brand of a vaccine, one can instead conduct a case-control study. If the potential adverse event is rare, cases can be identified through the health plan data, together with a set of matched controls. Chart review can then be conducted on this limited population to obtain more detailed information about each of the vaccines given, about the exact nature of the potential adverse event, or about various potential confounders.

There are also potential adverse events that are not fully captured in the electronic health plan data, such as neuropsychological performance or immune function. Such outcomes must be measured specifically for a research study, but that can obviously not be done for all members of the health plan. Depending on the outcome, there are at least three different ways to go about doing this:

- Select a random sample of children from the health plan; measure the outcome of interest on each one; and evaluate the relationship between the relevant component of the vaccine schedule and their outcome measurement values. This is the simplest approach.
- Select a non-random sample of children from the health plan, oversampling children on both end of the metric used in the study. For example, if the variable of interest is the timing of the first dose of the hepatitis B vaccine, children who received it long after birth would be oversampled. After that, proceed as above. This design will in many cases increase the statistical power. It is important that the probability of selection is unrelated to other health events.
- Select a non-random sample of children from the health plan, based on a health outcome that is present in the health plan data and that is correlated with the outcome of interest. The goal here is to get a larger variance in the health outcome being measured, thereby increasing the statistical power. It is important that the probability of selection is unrelated to any aspect of the vaccine schedule. After the study population has been selected, proceed as in the first scenario.

There is not necessarily a linear dose-response relationship between the metric of choice and the risk of a potential adverse event, but a linear function can be used in the regression model as a test for trend. Quadratic and other non-linear functions can then be explored and formally tested for statistical significance in order to get a better understanding about the dose-response curve.

9.   COMPARISON OF COMPLETE VACCINE SCHEDULES

Background: Some parents have consciously decided to follow a specific alternative vaccine schedule other than the one recommended by the CDC. Hence, there is a natural interest in comparing the safety of these complete schedules as discrete entities rather than through the various components of the schedule as discussed in the previous sections. For example, one may compare the number of potential adverse events in children that has followed the CDC recommended vaccine schedule, children that have followed one of Dr. Bob's recommended schedules (Sears 2007), and children that fall outside of either of these schedules.

Examples: In a pioneering study on vaccine schedules, Glanz et al. (2012) used a matched cohort design where children on the CDC recommended vaccine schedule were matched with children not on that schedule. It was then evaluated whether there were any differences in a few different outcomes: pertussis, upper respiratory infections, fever, sinusitis, outpatient physician visits, and hospitalizations. The data used was from the Vaccine Safety Datalink. In a companion study, Hambridge et al. (2012) used the same data to look at febrile seizures.

Data: As in the previous section, complete vaccination histories are required to classify children into alternative vaccine schedules, so electronic health plan data is again the best existing data set to use, as long as the length of enrollment is long enough for a sufficiently large number of children.

Methods: Obviously, very few children are one hundred percent compliant with any one particular schedule. Hence, the first challenge with these studies is to define some criteria of how divergent they can be from the schedule while still considered to be compliant. For example, one could require that all the recommended vaccines have been received and that average temporal divergence from the schedule is at most one month, over all the vaccines. Alternatively, one could require that the sum of the temporal divergences, taken over all vaccines, is at most, say, one year. If the criteria are too strict, the sample size will be too low and the statistical power will suffer. If the criteria are too wide, many of the children will not be true representatives of the schedule that they are set to represent, reducing the validity of the study.

Children with an incomplete vaccine history must be excluded from the study, together with children whose vaccine schedule is too divergent from either of the schedules being studied.

Once each child in the health plan has been classified according to the schedule, a variety of potential adverse events can be studied. A key difficulty here is to define the time period during which the events will be counted. The cleanest option is to make the vaccine schedule classification based on data up to a certain age and only consider potential adverse events that occur after this age. This ensures that there is

no bias if the adverse event studied causes subsequent changes in the vaccination schedule. It is not an ideal solution though, since we must either ignore the adverse events occurring early during the vaccination schedule, and/or the possible effect of later parts of the vaccination schedule. If the potential adverse event under study is such that there is little risk that its presence will change any aspect of the vaccination schedule, then one could include adverse events that occur before the end of the vaccine schedule considered, but such an approach is risky.

Since different children will have different lengths of follow-up, time-to-event data will best be analyzed using Cox proportional hazard survival analysis methods, adjusting for possible confounders. When two alternative schedules are compared, an alternative way to adjust for covariates is to use a matched cohort design (Glanz 2012), where each child with the recommended schedule is matched with a child with the alternative schedule, having the same age, gender, study site and calendar year. This is especially useful if additional health data is to be gathered from the children, which are not available in the health plan data sets, since it is typically infeasible to collect such data for all children in a health plan.

10.   ADDITIONAL METHODOLOGICAL ISSUES

Bias and Confounding: The observational study designs described above are, like all observational studies, prone to various sources of bias. The type of bias is often different for different designs, and it is hence often wise to use multiple study designs for the same question. Of special concern in these types of studies is confounding due to the innocent bystander effect. This is when an adverse event that is seemingly due to one component of the vaccine schedule under study is actually due to another component, which is correlated with the first one.

As a first step, it is natural to do a study considering one single component of the vaccine schedule, only adjusting for demographic variables. If a significant relationship is observed though, it is sometimes important to consider other aspects of the schedule as possible confounders. This is especially true for late onset events. This can be tackled in one of several ways. As a second step, additional studies can be conducted evaluating other components of the vaccine schedule. One way to do

this is to incorporate multiple components in the same regression model. For example, one regression model may include variables representing the age at which each vaccine was given, the total number of vaccines given, the total exposure to immune-stimulating content, and the total exposure to adjuvants. Such a design will provide information as to whether the component that is suspected of being the culprit still has a statistically significant relationship with the outcome after adjusting for other components of the schedule. It is important to note though, that with many different correlated components in the same schedule, none may be statistically significant after adjusting for all the others. This does not mean that risk of the adverse events does not depend on the vaccine schedule. It just means that it is not possible to determine which component is responsible, and that is important information.

Vaccine Specific versus General Components of Vaccine Schedules: The more general components of the vaccine schedule, described in Section 8, as well as the comparison between complete schedules in Section 9, are considerably more difficult to study than the more vaccine specific components described in Sections 3 to 7. There are several reasons for this.

First and foremost, there are many alternative vaccine schedules, and slightly different schedules have to be lumped together in the same comparison group. For the cumulative summary metrics, many different vaccination schedules will have the same value, such as, for example, the average age at vaccination. If one vaccine schedule is safer in terms of a specific outcome compared to an alternative vaccine schedule, but they are both have the same average age at vaccination, then the effect size will be attenuated and go undetected.

If a statistically significant risk difference is found, a second problem with these designs is that it can be hard to know which aspect of the schedule caused the excess or reduced risk. Is it the timing of one specific vaccine, is it an interaction between two or more vaccines, or is it something else? Hence, any statistically significant findings will have to be followed-up with studies concerning more specific vaccine schedule components.

A third issue is confounding. While confounding is present in all observational studies, it is a greater problem when studying complete vaccine schedules. For example, children for whom most of the vaccines are delayed, as compared to the recommended schedule, may be different in terms of both health care utilization and socio-economic factors. This may bias the results, and the bias may exist whether the delayers are deliberately following an alternative schedule or not. The same type of confounding can be present when looking at more specific components of the vaccine schedule, but it is likely to be less strong, as such components are likely to have more random and less systematic variability than a complete schedule. A way to intuitively see this is to note that whatever it is that causes a general parental tendency to delay vaccinations, that will likely be more correlated with the average timing of all vaccinations than with the timing of a single vaccination.

To date, there have been few comparative studies evaluating the safety of different vaccine schedules. For the above-mentioned reasons, it is suggested that initially, the majority of such studies focus on the most vaccine specific components of the vaccine schedule described in Sections 3 to 7, as well as the content defined components in Section 8. Information from such studies will greatly facilitate the understanding of subsequent studies evaluating the more general components discussed in Section 8 as well as the comparison of complete vaccines schedules described in Section 9.

Randomized Trials: Randomized trials are the gold standard for scientific studies, and pre-market phase III vaccine randomized trials play an important role in the evaluation of vaccine adverse events. Because of their limited sample size, rare adverse events may not be detected though. The utility of randomized trials is more limited in post-market vaccine safety studies. For rare but serious adverse events, the study needs to have a very large sample size to detect a potential problem. For example, if one of every 10,000 vaccinated children has a very serious adverse event, that is not something that can be detected in a randomized trial with 40,000 children in each arm, for a total of 80,000. To see this, suppose that there are 4 adverse events in the vaccinated arm and none in the control arm receiving the placebo. Under the null hypothesis, the probability of all 4 being in the vaccinated arm is $(1/2)^4$=0.0625, which is not statistically significant, and hence; we cannot conclude that it was the vaccine that caused the adverse events. So, for rare adverse events,

we need data with hundreds of thousands of vaccinated children, and for such sample sizes, randomized trials are prohibitively expensive.

For more common adverse events, randomized trials have a potential role to play in post-market vaccine safety studies. There is little reason to use them to evaluate the general safety of a particular vaccine, since that is already covered by the phase III trials, but it is a different matter for vaccine safety issues related to the schedule. Questions for which randomized trials may be used include the order in which different vaccines are given and the timing between doses of the same or different vaccines. Hypotheses about potential adverse events may come from phase III trials or from observational post-marketing studies using data from health plans or spontaneous reporting systems.

If and when a randomized trial is conducted, it is important to consider the effect on herd immunity. If the two arms differ by delaying one or more vaccines by at most a few weeks, it is not a major issue. If vaccination in one arm is delayed for a much longer time period, or not given at all, it may reduce herd immunity. This may put children that are not participating in the study at increased risk for the disease, and this can be especially serious for immunocompromised children for which a vaccination is contraindicated. To minimize the negative effect on herd immunity, such randomized trials should be spread out geographically, so that there are at most a few additional unvaccinated children in any given location. In that way, non-participating children will not be at an increased risk of the disease, and equally important, those children randomized to the delayed vaccination will still have some protection against the disease from herd immunity.

Cross-National Comparisons: Different countries have different recommended vaccine schedules, so it may seem natural to do cross-national studies to compare the safety of the schedules, in an ecological study design. Unfortunately, this is very difficult to do well, and generally not recommended. The problem is that the incidence of most diseases varies by geographical region for reasons other than their vaccine schedule, such as genetics, diet, physical exercise, or other environmental factors. Any such cross-national study may hence be heavily biased. This does not mean that one cannot do studies that include data from multiple countries or regions, as long as each one has a range of different exposures in each place. In such

studies, the geographical region can easily be adjusted for in the analysis, in order to take the differential disease incidences into account.

Time Trend Evaluations: Another ecological study design is to take a particular country or region and compare time trends in disease incidence with temporal changes in the vaccination rate or vaccination schedule. This is also not recommended. In addition to vaccinations, there are many other reasons why the reported disease incidence may increase or decrease over time, including changes in environmental risk factors and changes in health care practice and diagnosis. Hence, an apparent temporal correlation between increasing disease incidence and increasing vaccine could be completely spurious. It should be pointed out that the bias can also go in the other direction. Even if there is no temporal correlation between disease incidence and vaccinations, a true relationship can be hidden by a compensatory effect from an unknown confounder.

Near-Real Time Safety Surveillance: In what is called 'rapid cycle analysis', the Vaccine Safety Datalink project has pioneered near real-time vaccine safety surveillance (Lieu et al. 2007, Yih et al. 2011). For newly approved vaccines and selected adverse events, weekly data feeds are received from the health plans and analyzed using continuous sequential statistical methods (Kulldorff et al. 2011). If there are specific concerns regarding the newly revised vaccine schedule, such rapid cycle analysis can also be implemented for many of the study designs described above.

Disease Causing Complications versus Adverse Events: It should be noted that in these types of studies, it is not always clear what is an adverse event and what is not. For example, a child may have (i) a febrile seizure that was caused by one or more of the vaccines, (ii) a febrile seizure caused by a disease, where the disease was an adverse event caused by one or more of the vaccines, (iii) a febrile seizure caused by a disease, where the child got the disease because he or she was not immunized against it, or (iv) a febrile seizure caused by a disease for which vaccinated children has a lower risk even though they were not directly vaccinated against it. Hence, an excess risk of seizures due to a particular vaccine schedule could either be due to vaccines given at a certain time when the child is more

sensitive to adverse events or to vaccines not given at a certain time when the child needed the vaccine protection.

If the same type of health event is caused by the vaccine among one group of children, as an adverse event, and by the disease among the same or another group of children, as a complication, then the vaccine may be found to cause an excess number the adverse events in a vaccinated population, since the non-vaccinated children benefit from herd immunity. Hence, findings about the risk of individuals in a mostly vaccinated population cannot necessarily be generalized to the population level.

Vaccine Efficacy and Effectiveness: This paper only covers the study of potential vaccine adverse events. If a study does not find an excess risk, all is fine, and there is no need to worry about vaccine efficacy. On the other hand, if a true differential risk of adverse events is found with respect to some component of the vaccine schedule, vaccine efficacy and effectiveness must also be considered when contemplating a revised vaccine schedule. Some vaccines, such as MMR, have a different immune response depending on the age of the child, and the vaccine efficacy therefore depends on the vaccine schedule. The timing of a vaccination also influences the time period during which the child is protected from the disease and the herd immunity of the population at large. Herd immunity can also be affected if a parent refuses future vaccinations after his or her child has had a vaccine adverse event that could have been avoided with a different schedule. While outside the scope of this paper, all these factors must be considered in a joint cost-benefit analysis before revising the recommended vaccine schedule if and when there is a finding of a vaccine schedule dependent adverse event.

11.   CONCLUSION

The comparative safety evaluation of different vaccine schedules is a complex and multifaceted task, and all aspects of the vaccine schedule are currently under-studied with regards to potential adverse events. A number of different study designs and methods can be used to evaluate different components of the schedule. For all known and most potential adverse events, it is recommended that a wide

variety of vaccine schedule components are evaluated. To directly evaluate complete vaccine schedules is more difficult, probably less fruitful, but not impossible. Such studies are most useful when conducted in parallel with studies of specific components of the schedule. This is especially important when there is a significant adverse event finding, since it is otherwise impossible to know which of the many features of the complete schedule is actually causing the adverse events.

This paper should not be utilized as a cookbook where definite study designs and methods are obtained and used for different classes of problems in a black-box type approach. Each study is unique, depending on the vaccine(s) under study, the potential adverse event(s) of interest, the data used, and the scientific research question. All those aspects need to guide the methodology. The goal of this paper is simply to show that a wide variety of study designs and methods are available to study the comparative safety of different vaccine schedules, and the hope is that some of the proposed methods can serve as a starting point when thinking about the most suitable designs and statistical methods to use for different studies.

This paper does not present an exhaustive list of study designs and methods that can be used for the comparative evaluation of potential adverse events due to different childhood vaccination schedules. As more and more such studies are performed, additional designs and methods will surely be developed and used.

12.   REFERENCES

Aaby P, Jensen H, Gomes J, Fernandes M, Lisse IM. The introduction of diphtheria-tetanus-pertussis vaccine and child mortality in rural Guinea-Bissau: an observational study. International Journal of Epidemiology, 2004, 33:374-380.

Almenoff JS, DuMouchel W, Kindman LA, Yang X, Fram D. Disproportionality analysis using empirical Bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting. Pharmacoepidemiology and Drug Safety, 2003, 12:517-521.

Baggs J, Gee J, Lewis E, Fowler G, Benson P, Lieu T, Naleway A, Klein NP, Baxter R, Belongia E, Glanz J, Hambidge SJ, Jacobsen SJ, Jackson L, Nordin J, Weintraub E. The Vaccine Safety Datalink: a model for monitoring immunization safety. Pediatrics, 2011, 127:S45-53.

Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, De Freitas RM. A Bayesian Neural Network Method for Adverse Drug Reaction Signal Generation. European Journal of Clinical Pharmacology, 1998; 54: 315–321.

Chen RT, Glasser JW, Rhodes PH, Davis RL, Barlow WE, Thompson RS, Mullooly JP, Black SB, Shinefield HR, Vadheim CM, Marcy SM, Ward JI, Wise RP, Wassilak SG, Hadler SC. Vaccine Safety Datalink project: a new tool for improving vaccine safety monitoring in the United States. Pediatrics. 1997, 99:765-773.

DeStefano F and the Vaccine Safety Datalink Research Group. The Vaccine Safety Datalink project. Pharmacoepidemiology and Drug Safety, 2001;10:403-406.

DeStefano F, Weintraub E, Price C. Immunogenic stimulation from vaccines and risk of autism. Presentation, Vaccine Safety Datalink Annual Meeting, Denver, 2012.

DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. American Statistician 1999;53:177-90.

Eng PM, Mast TC, Loughlin J, Clifford CR, Wong J, Seeger JD. Incidence of intussusceptions among infants in a large commercially insured population in the United States. Pediatric Infectious Disease Journal, 2012, 31:287-291.

Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. Pharmacoepidemiology and Drug Safety, 2001, 10:483-6.

Fireman B, Lee J, Lewis N, Bembom O, van der Laan M, Baxter R. Influenza vaccination and mortality: differentiating vaccine effects from bias. American Journal of Epidemiology, 2009, 170:650-656.

Glanz J, et al. Alternative Immunization Schedules and the Risk of Pertussis. Presentation, Vaccine Safety Datalink Annual Meeting, Denver, 2012.

Hambidge SJ, Glanz JM, France EK, McClure D, Xu S, Yamasaki K, Jackson L, Mullooly JP, Zangwill KM, Marcy SM, Black SB, Lewis EM, Shinefield HR, Belongia E, Nordin J, Chen RT, Shay DK, Davis RL, DeStefano F. Safety of trivalent inactivated influenza vaccine in children 6 to 23 months old. Journal of the American Medical Association, 2006; 296:1990-1997.

Hambridge SJ, et al. Seizures in Children with Recommended versus Delayed Infant Immunization Schedules. Presentation, Vaccine Safety Datalink Annual Meeting, Denver, 2012.

Hsiao CJ, Hing E, Socey TC, Cai B. Electronic health record systems and intent to apply for meaningful use incentives among office-based physician practices: United States, 2001–2011. NCHS data brief, no 79. Hyattsville, MD: National Center for Health Statistics, 2011.

Iqbal S, DeStefano F. Immunological-stimulating Content of Vaccines and Neuro-physiological Outcomes in Children. Presentation, Vaccine Safety Datalink Annual Meeting, Denver, 2012.

Klein NP, Fireman B, Yih WK, Lewis E, Kulldorff M, Ray P, Baxter R, Hambidge S, Nordin J, Naleway A, Belongia EA, Lieu T, Baggs J, Weintraub E, for the Vaccine Safety Datalink. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. Pediatrics, 2010, 126, e1-8.

Kramarz P, France EK, Destefano F, Black SB, Shinefield H, Ward JI, Chang EJ, Chen RT, Shatin D, Hill J, Lieu T, Ogren JM. Population-based study of rotavirus vaccination and intussusception. Pediatric Infectious Disease Journal, 2001, 20:410-416.

Kulldorff M, Davis RL, Kolczak M, Lewis E, Lieu T, Platt R. A maximized sequential probability ratio test for drug and vaccine safety surveillance. Sequential Analysis, 2011, 30:58-78.

LaRussa PS, Edwards KM, Dekker CL, Klein NP, Halsey NA, Marchant C, Baxter R, Engler RJM, Kissner J, Slade BA. Understanding the role of human variation in vaccine adverse events: The Clinical Immunization Safety Assessment Network. Pediatrics, 2011, 127:S65-S73.

Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih WK, Yin R, Brown JS, Platt R. Real-time vaccine safety surveillance for the early detection of adverse events. Medical Care, 2007, 45:S89-95.

Leonardi M, Bromberg K, Baxter R, Gardner JL, Klopfer S, Nicholson O, Brockley M, Trammel J, Leamy V, Williams W, Kuter B, Schödel F. Immunogenicity and safety of MMRV and PCV-7 administered concomitantly in healthy children. Pediatrics, 2011, 128:e1387-94.

Murphy TV, Gargiullo PM, Massoudi MS, Nelson DB, Jumaan AO, Okoro CA, Zanardi LR, Setia S, Fair E, LeBaron CW, Wharton M, Livengood JR; Rotavirus Intussusception Investigation Team. Intussusception among infants given an oral rotavirus vaccine. New England Journal of Medicine, 2001, 344:564-572.

Nguyen M, Ball R, Midthun K, Lieu TA. The Food and Drug Administration's Post-Licensure Rapid Immunization Safety Monitoring program: strengthening the federal vaccine safety enterprise. Pharmacoepidemiology and Drug Safety, 2012, 21:S291-297.

Norén GN, Sundberg R, Bate A, Edwards IR. A statistical methodology for drug-drug interaction surveillance. Statistics in Medicine, 2008, 27:3057-3070.

Pittman PR. Aluminum-containing vaccine associated adverse events: role of route of administration and gender. Vaccine, 2002, 20:S48-50.

Price CS, Thompson WW, Goodson B, Weintraub ES, Croen LA, Hinrichsen VL, Marcy M, Robertson A, Eriksen E, Lewis E, Bernal P, Shay D, Davis RL, DeStefano F. Prenatal and infant exposure to thimerosal from vaccines and immunoglobulins and risk of autism. Pediatrics, 2010, 126:656-64.

Rothman KJ, Young-Xu Y, Arellano F. Age Dependence of the Relation between Reassortant Rotavirus Vaccine (RotaShield) and Intussusception. Journal of Infectious Diseases, 2006;193:898.

Rothman KJ, Lanes S, Sacks ST. The Reporting Odds Ratio and Its Advantages Over the Proportional Reporting Ratio. Pharmacoepidemiology and Drug Safety, 2004; 13:519–523.

Rowhani-Rahbar A, Lewis N, Fireman B, Baxter R, Klein N, et al. Effect of Age on the Risk of Seizure following Immunization with Measles-containing Vaccines. Presentation, Vaccine Safety Datalink Annual Meeting, Denver, 2012.

Simonsen L, Viboud , Elixhauser A, Taylor R, Kapikian AZ. More on RotaShield and Intussusception: The Role of Age at the Time of Vaccination. Journal of Infectious Diseases, 2005, 192:S36-S43.

Sears RW. The Vaccine Book: Making the Right Decision for Your Child. New York, NY: Little, Brown; 2007.

Thakrar BT, Grundschober SB, Doessegger L. Detecting signals of drug-drug interactions in a spontaneous reports database. British Journal of Clinical Pharmacology, 2007, 64:489-495.

Tse A, Tseng HF, Greene SK, Vellozzi C, Lee GM. Signal identification and evaluation for risk of febrile seizures in children following trivalent inactivated influenza vaccine in the Vaccine Safety Datalink Project, 2010-2011. Vaccine, 2012, 30:2024-2031.

Verium JE, Sodemann M, Biai S, Jakobsen M, Garly ML, Hedegaard K, Jensen H, Aaby P. Routine vaccinations associated with divergent effects on female and male mortality at the paediatric ward in Bissau, Guinea-Bissau. Vaccine, 23:1197-1204.

Yih WK, Kulldorff M, Fireman BH, Shui IM, Lewis EM, Klein NP, Baggs J, Weintraub ES, Belongia EA, Naleway A, Gee J, Platt R, Lieu TA. Active Surveillance for Adverse Events: The Experience of the Vaccine Safety Datalink Project. Pediatrics, 2011, 127:S54-64.