U.S. Advisory Committee on Immunization Practices (ACIP) Handbook for Developing Evidence-based Recommendations

Version 1.2

[November 1, 2013]

Faruque Ahmed, PhD
Centers for Disease Control and Prevention (CDC)
Atlanta, GA, USA

Table of Contents

Updates to Version 1.2: The following sections have been expanded – 4, 7, 8, 11, 14, 15, 16, 17.

Acknowledgments: This handbook for developing evidence-based recommendations for immunization builds on the GRADE handbook and work of the ACIP's Evidence Based Recommendations Work Group. The findings and conclusions in this handbook are those of the authors and do not necessarily represent the views of the CDC.

Suggested Citation: Ahmed F. U.S. Advisory Committee on Immunization Practices Handbook for Developing Evidence-based Recommendations. Version 1.2. Atlanta, GA: Centers for Disease Control and Prevention (CDC); 2013. Available from http://www.cdc.gov/vaccines/acip/recs/GRADE/about-grade.html#resources

01. Introduction

The U.S. Advisory Committee on Immunization Practices (ACIP) provides expert external advice and guidance to the Director of the Centers for Disease Control and Prevention (CDC) and the Secretary of the Department of Health and Human Services (HHS) on use of vaccines and related agents for control of vaccine-preventable disease in the U.S. civilian population. Information on the charter, structure, role, procedures, and membership of the ACIP are available at http://www.cdc.gov/vaccines/acip/committee/index.html. Resources and tools for implementing ACIP recommendations are available at http://www.cdc.gov/vaccines/.

The ACIP unanimously voted during its October 2010 meeting to adopt the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach for developing evidence-based recommendations. The purpose of this handbook is to provide guidance to the ACIP workgroups on using the GRADE approach for assessing the type or quality of evidence and for using that evidence to inform recommendations. Key factors for developing recommendations include the balance of benefits and harms, type or quality of evidence, values and preferences, and health economic analyses.

The ACIP recommendation categories are –

- Category A: Recommendation that applies to all persons in an age- or risk-based group.
- Category B: Recommendation for individual clinical decision making.
- No recommendation/unresolved issue.

Category A recommendations will be made for all persons in an age group or for all persons in a risk-based group. The suggested phrasing for category A recommendations include the words *recommend*, *recommend against*, *should*, and *should not*. Category B recommendations will indicate that clinical decisions should be made on an individual basis, i.e., category B recommendations do not apply to all members of a group, but are used in context of clinician-patient interaction to determine if vaccination may be appropriate for that patient. Phrasing for category B recommendations includes the words *may*, and *suggest against*. In some instances, it is possible that the ACIP may decide not to make a recommendation if additional information is needed.

The body of evidence is to be categorized into four types that represent a general hierarchy reflecting the confidence in the estimated effect of vaccination on health outcomes (benefits, harms):

1. Randomized controlled trials, or overwhelming evidence from observational studies.
2. Randomized controlled trials with important limitations, or exceptionally strong evidence from observational studies.
3. Observational studies, or randomized controlled trials with notable limitations.
4. Clinical experience and observations, observational studies with important limitations, or randomized controlled trials with several major limitations.

The process for evaluating a body of evidence and moving from evidence to recommendations includes the following steps:

- Formulating specific questions to be answered by a recommendation.
- Identifying all important outcomes for every question (benefits, harms).
- Judging the relative importance of outcomes.
- Summarizing all relevant evidence for important outcomes.
- Categorizing the type or quality of evidence for each outcome.
- Categorizing the overall type of evidence across outcomes.
- Assessing health economic data.
- Assessing the underlying values related to the management options and outcomes.
- Judging the balance of desirable and undesirable effects.
- Deciding on the recommendation category.
- Formulating a recommendation.

02. Formulating Questions

The GRADE system is intended to be applied to questions about alternate intervention strategies (e.g. vaccination vs. no vaccination). It is not intended to be applied to background questions or "good practice recommendations" where the recommendation is not actionable or the alternative is not credible or is a violation of basic standards of care (e.g. definition of influenza-like illness; use of antiseptic techniques for vaccination).

The scope of each recommendation should be defined, with each recommendation answering a focused healthcare question. Each question should indicate the **P**opulation, **I**ntervention, **C**omparison, and **O**utcomes (PICO) of interest. The questions are the starting point for formulating recommendations, and will drive the research direction (inclusion and exclusion criteria for the literature search) and determine the types of information that will be both searched for and assessed.

03. Choosing and Ranking Outcomes

Important outcomes that need to be considered in making a recommendation need to be identified. The purpose is to identify the outcomes that will be important or critical for making recommendations, and to identify the data that should be sought through evidence retrieval and synthesis. The outcomes are to be selected based on what is important, not what was measured in studies. An important outcome for which evidence is lacking should not be ignored and excluded from the evidence tables; the lack of evidence may influence the ultimate recommendation.

Surrogate outcomes should be considered only when evidence about health outcomes is lacking. When this is the case, both the health outcomes and the associated surrogates that must be used as substitutes should be specified. The surrogate themselves (e.g. immunogenicity) should not be listed as the measures of outcome.

ACIP workgroup members should make an initial list of possibly relevant outcomes, including both desirable and undesirable effects. Each member should be asked to score the importance of each outcome on a 1 to 9 scale using a modified Delphi process, where 7–9 indicates that the outcome is critical for a decision, 4–6 indicates that it is important but not critical, and 1–3 indicates that it is of limited importance. Guidance on using the modified Delphi process to rate importance is provided in *The RAND/UCLA Appropriateness Method User's Manual*, chapters 6–8 (available from http://www.rand.org/pubs/monograph_reports/MR1269.html). If a surrogate outcome is used, the importance of the corresponding health outcome rather than that of the surrogate outcome should be scored. The average score for each outcome can be used to determine its relative importance, though it is helpful to provide the range of results as well.

The importance of health outcomes is likely to vary within and across cultures or when considered from the perspective of the general population, patients, clinicians, or policy-makers. Workgroup members should decide what perspective they are taking. The perspective might vary by type of outcome (e.g. benefits, harms, cost-effectiveness).

The importance of health outcomes should be ranked before the evidence is reviewed to focus the evidence search and to clarify or resolve disagreements. However, after the evidence review, the ranking should be reassessed in light of the evidence review. It is possible that in some instances the importance of an outcome may only become known after the evidence has been reviewed. For instance:

- An outcome pertaining to a benefit may have been judged initially to be critical for making a recommendation, but it may no longer be considered to be critical if other benefits are evident; or
- A suspected adverse event may be initially considered to be critical, but if the evidence review shows that the adverse event is not causally associated with the intervention, it may be considered important but not critical.

Outcomes that are important or critical to decision making should be included in evidence tables whether or not information about them is available. Only outcomes considered *critical* are the primary factors influencing a recommendation and should be used to determine the *overall evidence type* supporting a recommendation.

04. Evidence Retrieval

A summary of all relevant research evidence is essential when developing a recommendation. Evidence should be sought relating to all important and critical outcomes. The most important type of evidence is that concerning the effect of interventions being considered in the recommendation. A recommendation should be based on the best available evidence related to the question that has been formulated.

If the body of evidence from randomized controlled trials (RCTs) for a question is rated as evidence type 1 (the highest level of evidence), it may not be necessary to evaluate observational studies. However, if evidence from RCTs is downgraded to evidence type 2, 3, or 4, evidence

from observational studies may provide complementary information that may prevent such a downgrade, or observational studies might provide a higher evidence level than RCTs.

Systematic methods should be used to identify the evidence. In contrast to narrative reviews, systematic methods address a specific question and apply a rigorous scientific approach to the selection, appraisal and synthesis of relevant studies. A systematic approach requires documentation of the search strategy used to identify all relevant published and unpublished studies and the eligibility criteria for the selection of studies. Systematic methods reduce the risk of selective citation and improve the reliability and accuracy of decisions. The Cochrane handbook provides guidance on searching for studies, including grey literature and unpublished studies (Chapter 6, Section 6.2, available at http://handbook.cochrane.org/).

An expert librarian should be consulted prior to conducting a search. He or she can help formulate a strategy that includes specific search terms and can assist in searches. This strategy should be documented and should specify:

- The details of the databases to be searched, and the search strategy to be applied to each database, e.g. PubMed; EMBASE; Cochrane Central Register of Controlled Trials (CENTRAL); databases of systematic reviews (Cochrane Library, etc.); databases of guidelines (US National Guideline Clearinghouse, etc.).
- The details of each strategy as actually performed, with search terms (key words and/or MESH terms), the date(s) on which the search was conducted and/or updated, and the publication dates of the literature covered.

Searches should be supplemented by reviewing references in the included studies and reviews, examining clinical trials registries maintained by the federal government (www.clinicaltrials.gov) and vaccine manufacturers, and consulting subject matter experts.

The criteria for including/excluding evidence identified by the search, and the reasons for including and excluding evidence, should be described (e.g. population characteristics, intervention, comparison, outcomes, study design, setting, language). The PRISMA Statement (www.prisma-statement.org) and AMSTAR instrument (http://amstar.ca/) include guidance on reporting the methods for evidence retrieval.

05. Type or Quality of Evidence

The evidence is to be assessed separately for each outcome. The body of evidence (not individual studies) is to be categorized into four types:

1) Randomized controlled trials, or overwhelming evidence from observational studies;
2) Randomized controlled trials with important limitations, or exceptionally strong evidence from observational studies;
3) Observational studies, or randomized controlled trials with notable limitations;
4) Clinical experience and observations, observational studies with important limitations, or randomized controlled trials with several major limitations.

The above evidence categories represent a general hierarchy that reflect confidence in the estimated effect of an intervention on health outcomes. Randomization reduces potential bias and confounding, and randomized controlled trials are considered the gold standard for assessing vaccine efficacy. However, observational studies may provide more relevant information than randomized trials in certain situations (e.g. to assess rare adverse events).

06. GRADE Criteria Determining Type or Quality of Evidence

The GRADE approach for assessing the type or quality of evidence involves consideration of several criteria in addition to the traditional assessment of risk of bias (i.e. study limitations), as shown in the Table below.

Table. GRADE criteria for assessing type or quality of evidence

<table>
<tr><th>Study design</th><th>Initial evidence Type[a]</th><th>Criteria for downgrading</th><th>Criteria for upgrading[b]</th><th>Final evidence type[a]</th></tr>
<tr><td>Randomized controlled trials</td><td>1</td><td rowspan="4">Risk of bias
-1 Serious
-2 Very serious

Inconsistency
-1 Serious
-2 Very serious

Indirectness
-1 Serious
-2 Very serious

Imprecision
-1 Serious
-2 Very serious

Publication bias
-1 Likely
-2 Very likely</td><td rowspan="4">Strength of association
+ 1 Large
+ 2 Very large

Dose response
+ 1 Evidence of a gradient

Opposing plausible residual confounding or bias
+ 1 Would reduce a demonstrated effect, or
+ 1 Would suggest an effect when results show no effect</td><td>1</td></tr>
<tr><td></td><td></td><td>2</td></tr>
<tr><td>Observational studies</td><td>3</td><td>3</td></tr>
<tr><td></td><td></td><td>4</td></tr>
</table>

[a]Evidence type:

1 = Randomized controlled trials (RCTs), or overwhelming evidence from observational studies.
2 = RCTs with important limitations, or exceptionally strong evidence from observational studies.
3 = Observational studies, or RCTs with notable limitations.
4 = Clinical experience and observations, observational studies with important limitations, or RCTs with several major limitations.

[b]Occasionally, other considerations that do not fit into the three criteria listed here may raise confidence in the findings and therefore warrant upgrading the evidence type.

Assessing the type or quality of the body of evidence for each outcome begins with the study design. Studies are classified into two types:

- Randomized controlled trials (RCTs);
- Observational studies. Examples include cohort studies, case-control studies, controlled before-after studies, interrupted time series studies, case series, case reports.

The PICO question needs to be taken into consideration when determining the classification of a study. For example, a study in which infants are randomized into two different vaccination schedules would be classified as an RCT if the question concerns which vaccination schedule is more effective, but it would be classified as an observational study with no control group if the comparison group of interest is infants who do not receive vaccination.

RCTs are initially classified as evidence type 1, and observational studies as evidence type 3. Five GRADE criteria are used for downgrading the evidence type: risk of bias; inconsistency; indirectness; imprecision; and publication bias. Three GRADE criteria are primarily used to upgrade the evidence type: strength of association; dose-response; and opposing plausible residual confounding or bias.

The GRADE criteria for upgrading or downgrading the evidence level may be additive. For example, when well-performed observational studies demonstrate both strength of association and dose-response, the evidence type may be upgraded by two levels to 1. Reviewers should categorize the final evidence type by considering an individual GRADE criterion in the context of strengths or limitations identified in any of the other GRADE criteria. For example, if limitations pertaining to the risk of bias and indirectness criteria are identified, but these limitations are not serious enough for moving down each of them, the evidence type may be downgraded by one level when limitations for both criteria are considered together. The GRADE criterion that played the biggest role in downgrading as well as all contributing factors should be specified.

07. Criteria for Downgrading Evidence Level

There are five GRADE criteria for assessing limitations that can lower the evidence level for randomized trials and observational studies: risk of bias; inconsistency; indirectness; imprecision; and publication bias.

7.1. Risk of Bias (Study Limitations)

Study limitations may bias the estimates of the effect of an intervention on health outcomes. The assessment of risk of bias should apply to studies contributing to results in the evidence tables, rather than to all studies that could potentially be included in the analysis. The criteria for evaluating study limitations or risk of bias (also referred to as internal validity) will depend on the study design. The number of studies is not a determining factor in determining risk of bias; a single well-conducted multicenter RCT may result in high confidence in the estimated effect of vaccination on health outcomes.

Randomized Control Trials

For RCTs, criteria established by the Cochrane Collaboration include: allocation sequence generation (selection bias); allocation sequence concealment (selection bias); blinding of participants and personnel (performance bias); blinding of outcome assessment (detection bias); incomplete outcome data (attrition bias); and selective outcome reporting (reporting bias) (ref: Cochrane handbook, Chapter 8, Section 8.5, http://www.cochrane-handbook.org/). Studies where study participants are allocated to intervention or control arms through quasi-randomization techniques (e.g. allocation by odd or even date of birth, date or day of admission, case record number, alternation/rotation) will automatically be downgraded because of risk of selection bias due to inadequate generation of a randomized sequence and ability of participants or investigators enrolling participants to foresee allocation. Blinding of outcome assessors is less important for the assessment of objective outcomes such as all-cause mortality, but crucial for subjective outcomes. Risk of bias can differ across outcomes (e.g. higher risk of bias for subjective outcomes compared to objective outcomes when outcome assessors are not blinded; different subsets of studies for safety vs. efficacy studies). For adverse events or non-inferiority studies, intention-to-treat analyses may not be appropriate (ref: Cochrane handbook, Chapter 16, Section 16.2.1). If any information needed for assessing risk of bias is not reported in a publication, one option is to contact study investigators. It may sometimes be possible to assess risk of bias from other reported information. For example, if information on allocation sequence concealment is not reported, data showing that the intervention and control groups are balanced at baseline may assuage concern regarding risk of bias. Assessing risk of bias due to incomplete outcome data involves considering the reasons for the missing data as well as the numbers missing (ref: Cochrane handbook, Chapter 8, Section 8.13). A relatively simple method for assessing the effect of missing data on risk of bias, which involves making plausible assumptions about the outcomes of persons with missing data, is described in the following articles:

- Akl EA, Johnston BC, Alonso-Coello P, et al. Addressing dichotomous data for participants excluded from trials: a guide for systematic reviewers. PLoS One, 2013;8(2):e57132.
- Ebrahim S, Akl EA, Mustafa RA, et al. Addressing continuous data for participants excluded from trial analysis: a guide for systematic reviewers. Journal of Clinical Epidemiology, 2013 Sep;66(9):1014-21.

For cluster RCTs, the Cochrane Collaboration's criteria for assessing risk of bias include assessment of the following: recruitment bias; baseline imbalance; loss of clusters; failure to account for clustering in the analysis; and comparability with individually randomized trials (ref: Cochrane handbook, Chapter 16, Section 16.3).

Observational Studies

Observational studies include cohort; case-control; controlled before-after; interrupted time series; case series; and case reports. For cohort and case-control studies, domains for evaluating risk of bias include: selection of the study groups; comparability of the groups; and ascertainment of either the outcome or exposure for cohort and case-control studies, respectively (ref: Newcastle-Ottawa Scale, http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm). For interrupted time series (e.g. population-based surveillance studies with multiple data points before and after

introduction of vaccination) and controlled before-after studies, evaluation criteria developed by the Cochrane Effective Practice and Organization of Care Group can be used (ref: EPOC Reviewer Tools, http://epoc.cochrane.org/epoc-author-resources). Case series and case reports are observational studies that investigate only persons exposed to the intervention. Source of control group results is implicit or unclear, thus, they will usually warrant classification as observational studies with important limitations. An explicit control group, however, may not always be necessary. For example, if vaccination of a large number of representative persons shows high seroconversion rates (referred to as case series or single group cohort study) AND if seroconversion rates in unvaccinated persons would be expected to be near zero, the absence of a concurrent control group may not be considered to be an important limitation. In a similar vein, if post-marketing phase 4 studies of a very large number of vaccine recipients show that the incidence of an adverse event is rare, and if one were confident that all events were detected, one would have high confidence that even if there were an association, the magnitude of absolute risk would be very small. The absolute risk of an adverse event may be more relevant for patients and populations than the relative risk.

Considerations on Study Limitations

Reviewers should consider the extent to which study limitations may bias the results. The risk of bias for each outcome is to be assessed using the following categories:

- No serious limitations (do not downgrade evidence type): most of the studies comprising the body of evidence have low risk of bias for all key criteria for evaluating study limitations.
- Serious limitations (downgrade one level): most of the studies have crucial limitation for one criterion or some limitations for multiple criteria that lower confidence in the estimated effect of vaccination on the outcome of interest.
- Very serious limitations (downgrade two levels): most of the studies have crucial limitation for one or more criteria that substantially lower confidence in the estimated effect.

When considering a body of evidence in which some studies have no serious limitations, some have serious limitations, and some have very serious limitations, it is not appropriate to automatically assign an average rating of serious limitations for the group of studies. When the risk of bias varies across studies, principles for determining whether to downgrade the evidence type for a group of studies include:

- Consider the extent to which each study contributes to the overall or pooled estimate of effect. Larger studies with many outcome events will contribute more;
- Assess whether the results differ for studies with low risk of bias and those with high risk of bias and consider focusing on studies with lower risk of bias if the results differ by risk of bias;
- Downgrade when there is substantial risk of bias across most of the studies;
- Consider limitations pertaining to the other GRADE criteria (if there are close calls regarding risk of bias with another GRADE criterion, consider downgrading the evidence level for at least one of the two GRADE criteria).

When close-call situations occur, this should be made explicit, and the reason for the ultimate classification should be stated.

7.2. Inconsistency

Inconsistency refers to an unexplained heterogeneity in the magnitude of the effect size (e.g. relative risk or odds ratio for binary outcomes; mean difference for continuous outcomes) across studies. If only one study is available, the inconsistency criterion is not applicable. Inconsistency can be assessed using point estimates, overlap of confidence intervals, and statistical tests of heterogeneity and I^2. The following indicate inconsistency:

- Point estimates vary widely across studies;
- Confidence intervals do not overlap or show minimal overlap;
- Statistical test for heterogeneity shows P-value of < 0.05;
- I^2 is large (I^2 around 50% is moderate and >75% is considerable).

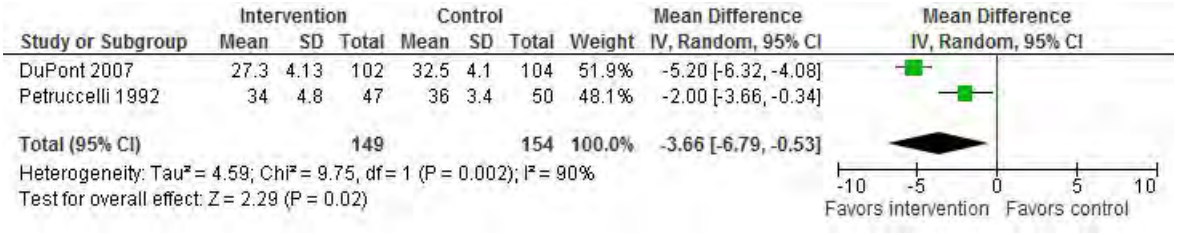
Inconsistency in results may arise from differences in:

- Populations (e.g. vaccines may have different relative effects in sicker populations);
- Interventions (e.g. different effects with different number of doses or comparators);
- Outcomes (e.g. duration of follow-up);
- Study methods (e.g. studies with higher and lower risk of bias).

When heterogeneity is large, but a plausible explanation can not be identified, the evidence level should be downgraded by one or two levels, depending on heterogeneity in the magnitude of effect. If inconsistency can be explained, estimates of effect should be presented separately for the stratification that explains the observed heterogeneity. If results differ by study methods, preference may be given to results of studies with a lower risk of bias. If results differ by population groups, different recommendations may be made for different groups.

For binary outcomes, inconsistency should be assessed using risk ratio or odds ratio. Risk difference should not be used to assess inconsistency, as risk difference is very sensitive to the baseline risk (i.e. risk in control group) and baseline risk can differ substantially between studies. The forest plot below shows three studies with consistent risk ratios for a binary outcome: point estimates are compatible across the studies; confidence intervals overlap; P-value for heterogeneity = 0.28; I^2 = 22%.

The forest plot below for a continuous outcome shows non-overlapping confidence intervals of the mean difference and a large I^2, but the differences are between small and large beneficial effects. The evidence level may or may not be downgraded for inconsistency. The decision for downgrading would depend on factors such as whether the uncertainty about the magnitude of benefit would influence the overall judgment about net benefit across all outcomes.



7.3. Indirectness

Evidence can be indirect in the following situations:

- The population that participated in studies may differ from the population of interest;
- The intervention that was evaluated may differ from the intervention of interest;
- The outcome that was assessed may differ from that of primary interest;
- The primary interest is head-to-head comparisons of vaccine A to vaccine B, but A was compared with C and B was compared with C.

Table. Examples of indirect evidence

Indirect	Question of Interest	Source of Indirectness
Population	Efficacy of vaccine in preventing disease in older persons with chronic health conditions	Studies are available for healthy persons, but not for the population of interest
Intervention	Efficacy of a new formulation of a vaccine in preventing disease	Studies of previous formulations of the vaccine provide indirect evidence bearing on the new vaccine
Comparator	Efficacy of vaccine A compared to vaccine B in preventing disease	Studies compared vaccine A to placebo and vaccine B to placebo, but studies comparing A to B are unavailable
Outcome	Prevention of disease	Increase in antibody titers following vaccination are reported, but there are no well-established standard correlates of protection

Indirectness can lower the evidence level by one or even two levels. To assess indirectness or applicability pertaining to the population, consider whether there are compelling reasons to think that differences in biological or social factors would result in substantial differences in the magnitude of the relative effect as opposed to differences in absolute effects that are influenced

by differences in the baseline incidence of disease across population groups. Because many interventions have similar relative effects across most population groups, assessment of directness across population groups should not be excessively stringent. On the other hand, studies using surrogate outcomes generally provide less direct evidence than those using health outcomes. Immunogenicity is a surrogate for disease incidence that would result in lowering of the evidence level unless there are well-established standard correlates of protection. Because evidence levels reflect the level of confidence in the estimated effect of vaccination in reducing disease, considerations of the feasibility of obtaining direct evidence should *not* influence the decision to downgrade the evidence level due to indirectness.

Complementary information from observational studies may be helpful to assess whether evidence from RCTs is direct enough. For example, evidence from RCTs conducted in men may not be downgraded when the population of interest includes both men and women if observational studies show that an intervention has the same effect in men and women. Further information on assessing indirectness is available in the following article: Schunemann HJ, Tugwell P, Reeves BC, et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. Research Synthesis Methods 2013;4:49-62.

7.4. Imprecision

Imprecision refers to studies having relatively few participants and few events that result in wide confidence intervals around the effect size.

For systematic reviews, the following indicate imprecision for an outcome:

- Total sample size across all studies for an outcome is lower than the calculated sample size for a single adequately powered study (online calculators are available for sample size calculations, e.g. http://www.stat.ubc.ca/~rollin/stats/ssize/b2.html); or
- 95% confidence interval (CI) of the pooled or best estimate of effect size includes both no effect AND appreciable benefit or appreciable harm (even if sample size is adequate). For binary outcomes, the suggested threshold for appreciable benefit or appreciable harm that should be considered for downgrading is a relative risk of 0.75 or 1.25, respectively (i.e. relative risk reduction or relative risk increase greater than 25%). For example, a relative risk with a 95% CI of 0.90 to 1.20 would be considered to be precise (1.20 is below the threshold of 1.25), whereas a relative risk with a 95% CI of 0.90 to 1.30 would be considered to be imprecise (1.30 is above the threshold of 1.25). When an outcome is rare, 95% CIs of relative effects may be very wide but 95% CIs of absolute effects may be narrow; in such situations, the evidence level may not be downgraded. For continuous outcomes, the threshold for appreciable benefit or appreciable harm refers to the difference in score in the outcome that are perceived as important.

For developing a guideline or recommendation, additional consideration should be given to whether the evidence is adequate to support a particular recommendation taking into account all outcomes together. The evidence level may be downgraded because of imprecision in the following situations:

- When the recommendation is *for* an intervention and
    - the 95% CI *includes* both no effect AND an effect that represent a benefit that would outweigh potential harms
    - the 95% CI *excludes* no effect but the lower confidence limit crosses a threshold below which, given potential harms, one would not recommend the intervention
- When the recommendation is *against* an intervention and
    - the 95% CI *includes* no effect AND an effect that represent a harm that, despite the benefits, would still be unacceptable
    - the 95% CI *excludes* no effect but the upper confidence limit crosses a threshold above which, given the benefits, one would recommend the intervention.

It should be noted that concerns about imprecision in evidence from RCTs may be alleviated if complementary information from observational studies support the results of RCTs (ref: Schunemann et al. Research Synthesis Methods, 2013;4:49-62).

More information on assessing imprecision is available in the following article: Guyatt G, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence - imprecision. J Clin Epidemiol 2011;64:1283-93.

7.5. Publication Bias

Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the selective publication of studies. Publication bias arises when investigators fail to publish studies, typically those that show no effect. Publication bias should be suspected if the available studies are uniformly small and funded by industry. A funnel plot of studies, with magnitude of the effect size (e.g. relative risk or odds ratio for a binary outcome) on the x-axis and variance (proxy for sample size) on the y-axis, can help assess publication bias. A funnel plot with asymmetrical distribution suggests publication bias.

08. Criteria for Upgrading Evidence Level

There are three primary criteria for moving up the evidence level: strength of association; dose-response gradient; and opposing plausible residual confounding or bias. On occasions, particular design features may warrant moving up the evidence level. For example, well-conducted observational studies showing that a vaccine reduces disease due to vaccine serotypes but not due to non-vaccine serotypes may increase confidence in the results if it is thought that the effect of possible biases because of unmeasured confounders would be similar for vaccine serotypes and non-vaccine serotypes.

Observational studies that have been downgraded for any reason (risk of bias, imprecision, inconsistency, indirectness, or publication bias) should not be moved up using the criteria for upgrading. RCTs that have been downgraded should generally not be moved up, except under certain circumstances. RCTs that have been downgraded because of quasi-randomization (e.g. allocation by day of week) may be moved up. RCTs that have been downgraded because of indirectness may be moved up in some instances.

8.1. Strength of Association

When methodologically strong observational studies show strong and consistent estimates of the magnitude of a treatment or exposure effect, one may be confident about the results. The stronger the association, the less likely it is that all of the apparent benefit or harm can be explained by residual confounding or bias, even though observational studies are likely to provide an overestimate of the true effect. The stronger the strength of association, the stronger becomes the evidence. The evidence level may be upgraded by one level if the relative risk from at least 2 studies is approximately >2 or <0.5, and it may be upgraded by two levels if the relative risk is approximately >5 or <0.2. For odds ratios, similar thresholds can be used for upgrading when the baseline risk is below 20%; when the baseline risk is higher, odds ratios do not approximate risk ratios, and higher thresholds for odds ratios may be appropriate. Both the point estimate and the confidence interval should be considered when upgrading; the evidence level should usually not be upgraded if the confidence interval overlaps substantially with the threshold.

Strength of Association	Effect Measure[a]	Evidence Level
Strong	Relative Risk approximately >2 or <0.5 (based on consistent evidence from at least 2 studies)	Move up 1 level
Very strong	Relative Risk approximately >5 or <0.2	Move up 2 levels

[a]Relative risks of 0.5 and 0.2 correspond to vaccine efficacies of 50% and 80%, respectively. Vaccine efficacy = (1 – Relative Risk) x 100.

Table. Example of very strong association: Rotashield vaccination and risk of intussusception

Finding	**Design[a] (# studies)**	**Risk of bias**	**Inconsis-tency**	**Indirect-ness**	**Impreci-sion**	**Other considera-tions**	**Evidence type**
Increased risk 3-14 days after vaccination	Observa-tional (2)	No serious	No serious	No serious	No serious	**Yes**[b]	1

[a]Included cohort and case-control studies available at the time the ACIP withdrew its recommendation for use of Rotashield vaccine; excluded ecological studies.
[b]Upgraded initial evidence type of 3 by two levels because relative risk of intussusception for vaccinated compared to unvaccinated infants is greater than 5 (*strength of association*).

Table. Example of strong association: MMRV vaccination and risk of febrile seizure after dose 1[a]

Finding	Design (# studies)	Risk of bias	Inconsis-tency	Indirect-ness	Impreci-sion	Other consider-ations	Evidence type
Increased risk 5-12 days after vaccination	Observ-ational (2)	No serious	No serious	No serious	No serious	**Yes**[b]	2
Decreased risk 13-30 days after vaccination	Observ-ational (2)	No serious	No serious	No serious	**Yes**[c]	None	4

[a]MMRV compared to separate injections of MMR and Varicella vaccination for children ages 12-23 months.
[b]Upgraded initial evidence type of 3 by one level because relative risk ~2 based on consistent evidence from two studies (*strength of association*).
[c]Downgraded initial evidence type of 3 by one level because of *imprecision*. One study indicated a decrease but not significant, one study found no association (because the confidence intervals overlapped, the results were not considered to be heterogeneous).

8.2. Opposing Plausible Residual Confounding or Bias

All plausible residual confounding or bias may on occasion be working to reduce the demonstrated effect (or increase the effect if no effect was observed). For example, if a vaccine is suspected of being associated with an adverse event, and the publicity results in increased spontaneous reporting of the adverse event among vaccinated persons compared to that in unvaccinated persons, yet epidemiological studies find no association, the evidence level for the lack of association can be upgraded.

8.3. Dose-response Gradient

The presence of a dose-response gradient may increase confidence in the findings of observational studies and thereby increase the evidence level. Examples include greater vaccine efficacy with increasing number of doses, and declining disease with increasing population vaccination rates.

09. Indirect Evidence

Indirect evidence may help facilitate decision making. For example, for the development of the 2011 ACIP recommendation for routine use of quadrivalent HPV vaccine in males, data on the efficacy of HPV vaccine in preventing anal intraepithelial neoplasia (AIN) 2/3 – a surrogate for anal cancer – were available for the MSM (men who have sex with men) population but not for

the general population. The evidence type for the outcome anal cancer was downgraded by one level for indirectness because of the use of a surrogate, but it was not downgraded further for differences in population characteristics because there was no reason to suspect that the efficacy in preventing AIN 2/3 in the general population would differ from that observed in the MSM population.

10. Overall Evidence Type

The overall evidence type combines evidence across many outcomes considered critical for a recommendation. When the body of evidence for an outcome includes both RCTs and observational studies, the study design that provides higher evidence level should be selected for that outcome when determining the overall evidence type. The overall evidence type is generally the lowest evidence type of the critical outcomes. An exception to this guidance is when the evidence type differs across critical outcomes but the association is similar in direction. For example, if data from well-conducted RCTs (i.e. the highest level of evidence) show that a vaccine reduces incidence of disease and hospitalization, and data from ecological studies indicate protective effect due to herd immunity, the overall evidence type may be categorized as RCTs even if herd immunity is considered to be a critical outcome.

11. Pooling Effect Estimates

Summary effect estimates across studies can be generated using meta-analysis software, by either the fixed-effect model or the random-effects model. The fixed-effect model assumes that included studies are functionally identical. The random-effects model is appropriate if the true effect size could vary across studies because of functional differences in population, intervention, comparison, or outcome definitions that could have affected the results. The studies, however, should have enough in common for the pooled estimate to be meaningful.

Summaries of binary outcomes (events, total) or continuous outcomes (mean, standard deviation, total) for each study group can be entered into a meta-analysis software such as RevMan (available at http://ims.cochrane.org/revman). It is also possible to enter effect estimates and standard errors (or confidence intervals) directly into RevMan using the 'Generic inverse variance' method for the random-effects or fixed-effect model. The inverse variance method, for example, enables pooling of adjusted effect estimates (e.g. adjusted odds ratios) for observational studies (ref: Cochrane Handbook, Chapter 9, Section 9.4.3).

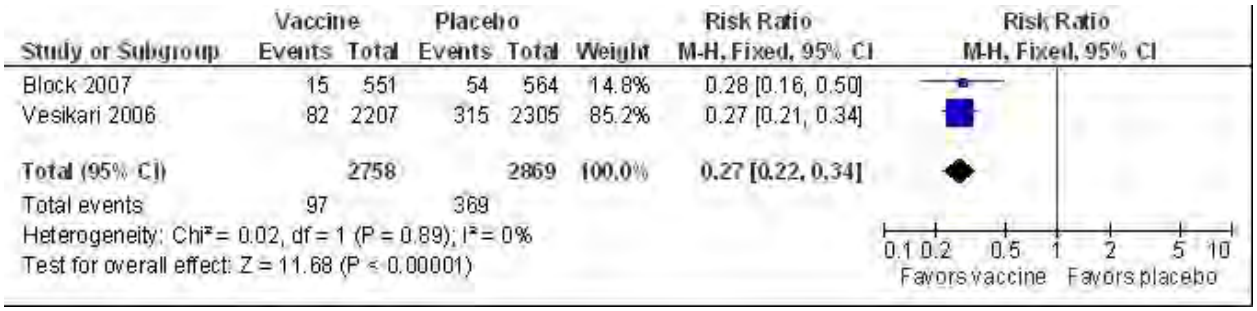
Studies that used two independent groups, matched groups, and clustered groups may be included in the same meta-analysis. Effect size and variance for each study is computed, taking into consideration the study design.

11.1 Binary Outcomes

Meta-analysis should be conducted using relative measures of effect (e.g. relative risk, odds ratio). It should be noted that using *pooled risk difference* estimates from meta-analyses may be

misleading, because risk difference is very sensitive to the baseline risk, and baseline risk can differ substantially between studies. However, when baseline risk is very low, confidence intervals computed from meta-analysis of relative measures of effect may be misleading, and direct computation of pooled risk difference using meta-analysis is preferable.

An example of a forest plot showing pooled relative risk and 95% confidence intervals (CIs) generated using RevMan software is shown below:



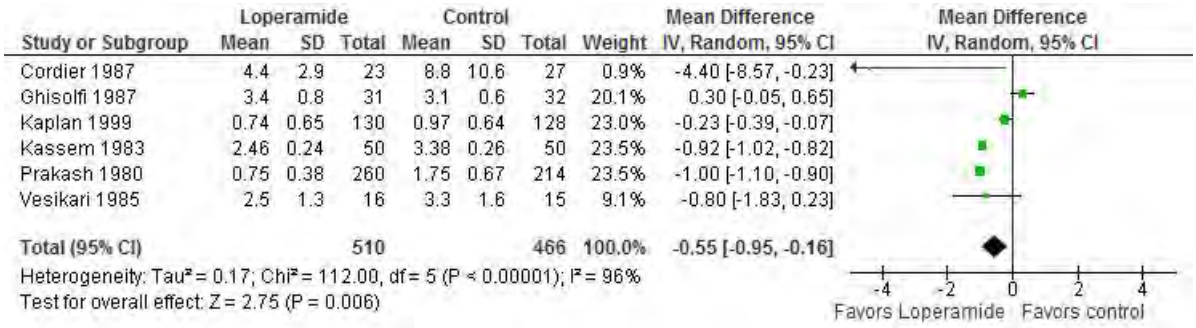
GRADEPro software can be used to generate risk difference estimates along with the confidence intervals for a range of baseline risks (see Appendix 3). Data on assumed incidence in controls and point estimate and 95% confidence interval of the pooled relative risk need to be entered into GRADEPro (or can be directly imported from RevMan) to generate the output. The assumed incidence in controls can be estimated from the control arms of studies (e.g. median incidence) or from other studies (e.g. surveillance studies). For example, if RCTs have assessed the effectiveness of an intervention in a population that differs from the population of interest, baseline risks in the population of interest derived from observational studies may be entered into GRADEPro to compute risk difference estimates. The formulae for computing risk differences or Number Needed to Treat (NNT) from the results of meta-analyses of risk ratios or odds ratios are given in the Cochrane handbook, Chapter 12, Section 12.5.4. The Cochrane handbook provides formulae for converting odds ratios to risk ratios (Chapter 12, Section 12.5.4.4). Additional information is available in the following article: Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines 12. Preparing Summary of Findings tables – binary outcomes. J Clin Epidemiol. 2012;66(2):158-72.

11.2 Continuous Outcomes

For continuous outcomes, raw mean difference, standardized mean difference, or response ratio can be computed. Raw mean difference is appropriate if the measurement scale is intuitively meaningful (e.g. duration of diarrhea). If the measurement scale is not intuitively meaningful or if different studies use different scales for measuring an outcome, standardized mean difference should be used. Standardizing the mean difference in each study by dividing by the study's standard deviation results in an outcome measure that is comparable across studies.

An example of a forest plot using studies cited in a systematic review by Li et al (PLoS Medicine, 2007) on Loperamide therapy for acute diarrhea in children for the outcome duration of diarrhea in days is shown below. The mean, standard deviation (SD), and sample size for the outcome in the Loperamide and control groups for each study were entered into RevMan. The pooled raw

mean difference is 0.55 days lower in the Loperamide group compared to the control group (95% CI: 0.95 days lower to 0.16 days lower).



In situations where an outcome is reported as continuous data in some studies and as binary data in other studies, methods for combining continuous and binary data are provided in the Cochrane handbook, Chapter 9, Section 9.4.6.

11.3 Geometric Mean Titers (Skewed Continuous Outcome)

For geometric mean titers (GMTs), the data for each study need to be converted to a log scale for computing pooled estimates. The pooled results are then converted back into the original metric by taking their exponentials (anti-logs).

An example of an immunogenicity meta-analysis using published randomized trials cited in a systematic review by Banzhoff et al (Gerontology 2003;49:177-84) is shown below. Standard deviations (SDs) have been derived from the reported 95% confidence intervals (CIs); for studies for which CIs were not reported, SDs have been imputed (ref: Cochrane handbook, Chapter 7, Section 7.7.3.2).

Table. Post-immunization GMT on Day 28: Data extracted from published articles[a]

	Adjuvanted vaccine		Non-adjuvanted vaccine	
Study	n	GMT (95% CI)	n	GMT (95% CI)
Donato 1999	94	137 (115, 162)	98	84 (71, 99)
Gasparini 2001	192	102 (92, 114)	99	70 (59, 83)
Baldo 2001	99	75.6 (58.4, 92.9)	93	66.4 (54.9, 77.9)
Martin 1997	277	111	186	76
Menegon 1999	96	54.5	98	49.0
Minutello 1999	46	115	46	74

[a]95% CIs were not reported for the Martin, Menegon, and Minutello studies.

Table. Post-immunization GMT: Extracted data transformed to $\log_e$ scale

Study	Adjuvanted vaccine		Non-adjuvanted vaccine	
	n	Log_e GMT (Log_e 95% CI)	n	Log_e GMT (Log_e 95% CI)
Donato 1999	94	4.92 (4.74, 5.09)	98	4.43 (4.26, 4.60)
Gasparini 2001	192	4.62 (4.52, 4.74)	99	4.25 (4.08, 4.42)
Baldo 2001	99	4.33 (4.07, 4.53)	93	4.20 (4.01, 4.36)
Martin 1997	277	4.71	186	4.33
Menegon 1999	96	4.00	98	3.89
Minutello 1999	46	4.74	46	4.30

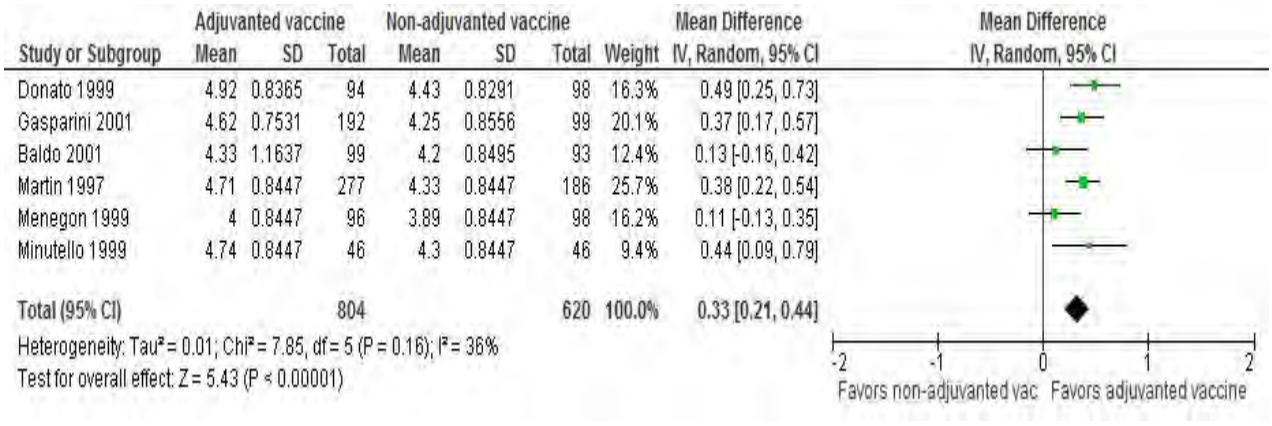
Table. Post-immunization GMT: Standard Deviations derived from 95% confidence intervals[a]

Study	Adjuvanted vaccine		Non-adjuvanted vaccine	
	n	Log_e GMT (Log_e SD)	n	Log_e GMT (Log_e SD)
Donato 1999	94	4.92 (0.8365)	98	4.43 (0.8291)
Gasparini 2001	192	4.62 (0.7531)	99	4.25 (0.8556)
Baldo 2001	99	4.33 (1.1637)	93	4.20 (0.8495)
Martin 1997	277	4.71 (0.8447)[b]	186	4.33 (0.8447)[b]
Menegon 1999	96	4.00 (0.8447)[b]	98	3.89 (0.8447)[b]
Minutello 1999	46	4.74 (0.8447)[b]	46	4.30 (0.8447)[b]

[a]Standard deviation = Standard error x Square root of n; where Standard error = (Upper confidence limit – Lower confidence limit) ÷ t value. [t value for a 95% confidence interval can be obtained by typing **=tinv(1-0.95,n-1)** in Excel, where n represents sample size)]. The Calculator in RevMan can be used to derive standard deviations from 95% confidence intervals.

[b]Log_e SD was imputed by using the average $\log_e$ SD of non-adjuvanted vaccines from the Donato, Gasparini, and Baldo studies.

Figure. Forest plot of post-immunization log-transformed GMT data



Pooled GMT Ratio, adjuvanted vaccine vs. non-adjuvanted vaccine = $e^{0.33}$ = 1.39
95% CI of pooled GMT Ratio = $e^{0.21}$, $e^{0.44}$ = 1.23, 1.55

Note: A sensitivity analysis where the extreme case $\log_e$ SD of 1.1637 was imputed instead of the average $\log_e$ SD resulted in a pooled GMT Ratio of 1.40 (95% CI: 1.25, 1.57).

12. Evidence Tables

The number of studies and participants, relative and absolute effects, and type or quality of evidence should be presented in tabular format to inform the development of recommendations. Each judgment pertaining to categorizing the evidence should be made explicit in order to increase the transparency of the process. Evidence tables should include footnotes explaining the reasons for upgrading or downgrading the evidence level.

Data on benefits and harms should be summarized across studies so that evidence tables have one row for each outcome. If there are both randomized trials and observational studies for a given outcome, the quality of evidence for each type of design should be presented.

Examples of evidence tables are provided in Appendix 1 (pages 36-38) and in the following article: Ahmed F, Lindley MC, Allred N, Weinbaum CM, Grohskopf L. Effect of influenza vaccination of health care personnel on morbidity and mortality among patients: Systematic review and grading of evidence. Clinical Infectious Diseases 2013; doi: 10.1093/cid/cit580.

13. Health Economic Analyses

Health economic evaluations take into account costs and health outcomes of an intervention in relation to its comparator. Effectiveness measures can be natural units (e.g. disease episodes or deaths prevented), quality-adjusted life years (QALYs), or can be expressed in monetary terms. Health economic analyses often use modelling in order to combine evidence from different sources and to extrapolate from the limited time-horizons of existing studies on health outcomes. The methodology described above for categorizing the type of evidence is not intended to be applied to health economic analyses based on modelling. Presentation of health economic data should be undertaken using the guidelines in the document titled *Guidance for Health Economic Studies Presented to the ACIP* (ref: MMWR 2008;57(5):125-6; http://www.cdc.gov/vaccines/recs/acip/economic-studies.htm).

14. Values and Preferences

Values and preferences can be described as the relative importance of outcomes related to benefits, harms, and costs. Ethical considerations can be considered under the rubric of values and preferences.

There will always be advantages and disadvantages of alternative management strategies, and individuals will always have to make a trade-off between them. Therefore, the way a workgroup values particular benefits and risks can be decisive to any recommendation. The values used in making recommendations should reflect those of the people affected (e.g. general population, patients, clinicians, policy-makers). While it is ideal to obtain values and preference estimates

from representative population-based studies, such studies may not be available. If there is a paucity of published information on values, values of workgroup members may be used as a proxy. The ranked outcomes (see Section 03: Choosing and Ranking Outcomes) provide an indication of values of workgroup members.

Methods for assessing values and preferences include discrete choice experiment, standard gamble, time trade off, and willingness to pay. Examples of articles on values and preferences include:

- Kuppermann M, Nease RF, Ackerson LM, Black SB, Shinefield HR, Lieu TA. Parents' preferences for outcomes associated with childhood vaccinations. Pediatr Infect Dis J 2000;19:129-33.
- Gidengil C, Lieu TA, Payne K, Rusinak D, Messonnier M, Prosser LA. Parental and societal values for the risks and benefits of childhood combination vaccines. Vaccine 2012;30:3445-52.
- Prosser LA, Ray T, O'Brien M, Kleinman K, Santoli J, Lieu TA. Preferences and willingness to pay for health states prevented by pneumococcal conjugate vaccine. Pediatrics 2004;113(2):283-90.
- Prosser LA, Payne K, Rusinak D, Shi P, Uyeki T, Messonnier M. Valuing health across the lifespan: Health state preferences for seasonal influenza illnesses in patients of different ages. Value in Health;2011;14:135-43.
- Prosser LA, Bridges CB, Uyeki TM, et al. Values for preventing influenza-related morbidity and vaccine adverse events in children. Health and Quality of Life Outcomes 2005;3:18. http://www.hqlo.com/content/3/1/18.
- Lavelle TA, Meltzer M, Gebremariam A, Lamarand K, Fiore AE, Prosser LA. Community-based values for 2009 Pandemic Influenza A H1N1 illnesses and vaccination-related adverse events. PLOS ONE 6(12):e27777.
- de Bekker-Grob EW, Hofman R, Donkers B, et al. Girls' preferences for HPV vaccination: A discrete choice experiment. Vaccine 2010;28:6692-7.
- Brown DS, Johnson FR, Poulos C, Messonnier ML. Mothers' preferences and willingness to pay for vaccinating daughters against human papillomavirus. Vaccine 2010;28:1702-8.
- Lieu TA, Ortega-Sanchez I, Ray T, et al. Community and patient values for preventing herpes zoster. Pharmacoeconomics 2008;26(3):235-49.
- Lee GM, Salomon JA, Gay C, Hammitt JK. Preferences for health outcomes associated with Group A Streptococcal disease and vaccination. Health and Quality of Life Outcomes 2010;8:28. http://www.hqlo.com/content/8/1/28.
- Lee GM, Salomon JA, LeBaron CW, Lieu TA. Health-state valuations for pertussis: methods for valuing short-term health states. Health and Quality of Life Outcomes 2005;3:17. http://www.hqlo.com/content/3/1/17.
- Kramer MS, MacLellan A, Ciampi A, Etezadi-Amoli J, Leduc DG. Parents' vs physicians' utilities (values) for clinical outcomes in potentially bacteremic children. J Clin Epidemiol 1990;43(12):1319-25.
- Bennett JE, Sumner W, Downs SM, Jaffe DM. Parents' utilities for outcomes of occult bacteremia. Arch Pediatr Adoles Med 2000;154:43-8.

- O'Meara JJ, McNutt RA, Evans AT, Moore SW, Downs SM. A decision analysis of streptokinase plus heparin as compared with heparin alone for deep-vein thrombosis. N Engl J Med 1994;330:1864-9.

15. Balance of Desirable and Undesirable Effects

Assessing the balance of benefits and harms for formulating recommendations involves considering the following factors:

- Relative importance of outcomes (i.e. relative values that patients and other stakeholders place on benefits and harms);
- Baseline risks of outcomes (i.e. incidence of outcomes in the absence of vaccination);
- Magnitude of relative risk;
- Magnitude of risk difference;
- Precision (i.e. 95% confidence interval) of relative risk and risk difference estimates.

The considerations for assessing the balance of desirable and undesirable effects include: point estimates of the effects of vaccination on all important and critical outcomes (e.g. relative risk, risk difference, NNT); and values and relative preferences attributed to these outcomes. Baseline risk of disease plays a role in determining the balance, since risk difference (and NNT) is a function of baseline risk and relative risk. In general, the higher the baseline risk, the greater is the magnitude of benefit.

16. Baseline Risk Estimates

The GRADE criteria for assessing the type or quality of evidence regarding the effect of an intervention (risk of bias, indirectness, imprecision, inconsistency, publication bias) may be useful for assessing the quality of evidence of baseline risks of outcomes (ref: Spencer FA et al. BMJ 2012;345:e7401, http://www.bmj.com/content/345/bmj.e7401). However, grading baseline risks may make the systematic review process overwhelmingly complex. Presenting the effect of a plausible range of baseline risk estimates on risk difference estimates may be sufficient to address possible concerns regarding the quality of baseline risk estimates. However, if necessary, uncertainty about baseline risk estimates can be classified under "indirectness" for assessing the quality of evidence regarding the effect of an intervention.

17. Formulating Recommendations

17.1 Recommendations Categories

Category A recommendations apply to all persons in an age group (e.g. routine recommendation) or to all persons in a specified risk group. Category B recommendations do not apply to everyone, but in the context of a clinician-patient interaction, vaccination may be found to be appropriate for a person. In some instances, the ACIP may decide not to make a recommendation if additional information is needed.

17.2 Factors Determining the Recommendation Category

The recommendation category reflects the balance between the desirable effects of adherence to the recommendation (benefits, savings) and the undesirable effects (harms, costs). The following table provides a brief explanation of the key factors for developing recommendations.

Key Factors	Explanation
Balance between benefits and harms	The larger the difference between the benefits and harms, the more likely is a category A recommendation warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely is a category B recommendation warranted.
Evidence type	The higher the evidence quality, the more likely is a category A recommendation.
Values and preferences	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely is a category B recommendation warranted.
Health economic analyses	The lower the cost-effectiveness of vaccination, the less likely is a category A recommendation warranted.

A *category A* recommendation is one for which the desirable effects clearly outweigh the undesirable effects (recommendation for) or that the undesirable effects clearly outweigh the desirable effects (recommendation against).

A hypothetical example of key factors that can result in a *category* A recommendation is:

Key Factors	Comments
Balance between benefits and harms	There is considerable benefit and little expected harm. Benefits are valued much higher than expected minor harms.
Evidence type	2
Values and preferences	Not a lot of variability in values and preferences in the target population.
Health economic analyses	Cost-effectiveness is high.

Key factors that can lead to *category B* recommendations include:

- Small benefits (e.g. small relative or absolute effects, low baseline risk);
- Lower confidence in the estimated effect of vaccination on health outcomes (e.g. studies with major limitations);
- Variation or uncertainty in how different individuals value the outcomes;
- Lower cost-effectiveness or lack of adequate data on cost-effectiveness.

The following article provides additional guidance on determining the recommendation category: Schunemann HJ et al. Grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. Am J Respir Crit Care Med, 2006;174:605-14.

17.3 Wording of Recommendations

The language used in the recommendation should be clear and direct, indicating an unambiguous action. The preferred wordings for category A recommendations are "recommend," "recommend against," "should," and "should not," and that for category B recommendations are words like "may" and "suggest against."

18. Presenting Recommendations

The recommendation is to be followed by the recommendation category and evidence type in parentheses. The key thought process behind the recommendation should be summarized.

Example:
Recommendation: ACIP recommends universal vaccination of U.S. infants with three doses of rotavirus vaccine administered orally at ages 2, 4, and 6 months (recommendation category: A, evidence type: 1).

Remarks: Nearly every child in the U.S. is infected with rotavirus by age 5 years, resulting in approximately 410,000 physician visits, 205,000–272,000 emergency department visits, and 55,000–70,000 hospitalizations each year. Vaccination reduces severe rotavirus diarrhea. The benefits are substantial compared to potential harms.

Additional guidance on presenting recommendations is included in the Appraisal of Guidelines for Research and Evaluation (AGREE II) instrument (available at www.agreetrust.org).

19. Further Research Needs and Priorities

A recommendation should summarize the available evidence as the basis of the recommendations. It is also useful to highlight where the available evidence regarding the effect of an intervention may be insufficient or inadequate. ACIP workgroups should identify research needs and, if appropriate, prioritize them. In formulating research needs, workgroups should be as specific as possible about what is needed and why. One format is EPICOT (ref: Brown P et al. BMJ 2006;333:804-6):

- E Evidence: What is the current state of the evidence (e.g. there is only one small observational study in older adults).
- P Population: What is the population of interest (e.g. older adults with chronic conditions).
- I Intervention: What are the interventions of interest (e.g. vaccination).
- C Comparison: What are the comparisons of interest (e.g. comparison to another vaccine).
- O Outcome: What are the outcomes of interest (e.g. disease, hospitalization, death, adverse events).
- T Time stamp: Date of literature search or recommendation (e.g. January 2012).

20. References

1. Ahmed F, Temte JL, Campos-Outcalt D, Schünemann HJ; for the ACIP Evidence Based Recommendations Work Group. Methods for developing evidence-based recommendations by the Advisory Committee on Immunization Practices (ACIP) of the U.S. Centers for Disease Control and Prevention (CDC). Vaccine. 2011;29:9171-6.

2. Schünemann H, Brozek J, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendation. Version 3.2 [updated March 2009]. The GRADE Working Group, 2009. Available from http://ims.cochrane.org/gradepro.

3. Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knotterus A. GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. J Clin Epidemiol. 2011;64(4):380-2.

4. Guyatt G, Oxman AD, Akl E, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, Debeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables**.** J Clin Epidemiol. 2011;64(4):383-94.

5. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, Alderson P, Glasziou P, Falck-Ytter Y, Schunemann HJ. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011;64(4):395-400.

6. Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol. 2011;64(4):401-6.

7. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J, Schünemann HJ.. GRADE guidelines: 4. Rating the quality of evidence – study limitations (risk of bias). J Clin Epidemiol. 2011;64(4):407-15.

8. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, Alonso-Coello P, Djulbegovic B, Atkins D, Falck-Ytter Y, Williams JW Jr, Meerpohl J, Norris SL, Akl EA, Schünemann HJ. GRADE guidelines: 5. Rating the quality of evidence – publication bias. J Clin Epidemiol. 2011;64(12):1277-82.

9. Guyatt G, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux P, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines: 6. Rating the quality of evidence - imprecision. J Clin Epidemiol. 2011;64(12):1283-93.

10. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Glasziou P, Jaeschke R, Akl EA, Norris S, Vist G, Dahm P, Shukla VK, Higgins J, Falck-Ytter Y, Schünemann HJ; The GRADE Working Group. GRADE guidelines: 7. Rating the quality of evidence - inconsistency. J Clin Epidemiol. 2011;64(12):1294-1302.

11. Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, Alonso-Coello P, Falck-Ytter Y, Jaeschke R, Vist G, Akl EA, Post PN, Norris S, Meerpohl J, Shukla VK, Nasser M, Schünemann HJ; The GRADE Working Group. GRADE guidelines: 8. Rating the quality of evidence - indirectness. J Clin Epidemiol. 2011;64(12):1303-10.

12. Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, Atkins D, Kunz R, Brozek J, Montori V, Jaeschke R, Rind D, Dahm P, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Murad MH, Schünemann HJ; The GRADE Working Group. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol. 2011;64(12):1311-16.

13. Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, Atkins D, Kunz R, Montori V, Jaeschke R, Rind D, Dahm P, Akl EA, Meerpohl J, Vist G, Berliner E, Norris S, Falck-Ytter Y, Schünemann HJ. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol. 2012;66:151-7.

14. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, Brozek J, Norris S, Meerpohl J, Djulbecovic B, Alonso-Coello P, Post PN, Busse JW, Glasziou P, Christensen R, Schünemann HJ. GRADE guidelines 12. Preparing Summary of Findings tables-binary outcomes. J Clin Epidemiol. 2012;66(2):158-72.

15. Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, Johnston BC, Karanicolas P, Akl EA, Vist G, Kunz R, Brozek J, Kupper LL, Martin SL, Meerpohl JJ, Alonso-Coello P, Christensen R, Schunemann HJ. GRADE guidelines 13. Preparing Summary of Findings tables-continuous outcomes. J Clin Epidemiology 2013;66(2):173-83.

16. Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, Nasser M, Meerpohl J, Post PN, Kunz R, Brozek J, Vist G, Rind D, Akl EA, Schünemann HJ. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol. 2013;66(7):719-25.

17. Andrews JC, Schünemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, Rind D, Montori VM, Brito JP, Norris S, Elbarbary M, Post P, Nasser M, Shukla V, Jaeschke R, Brozek J, Djulbegovic B, Guyatt G. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol. 2013;66(7):726-35.

18. Schunemann HJ, Tugwell P, Reeves BC, Akl EA, Santesso N, Spencer FA, Shea B, Wells G, Helfand M. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. Research Synthesis Methods 2013;4:49-62.

19. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester, UK: John Wiley and Sons; 2009.

Appendix 1. Example of Applying GRADE Framework

In this example, the GRADE framework is applied in a retrospective manner to the 2006 ACIP recommendation for use of pentavalent rotavirus vaccine (MMWR 2006; Vol 55; RR-12:1-13). Studies that were available at the time of the 2006 ACIP recommendation are used.

Background

In the United States, rotavirus infection is responsible for approximately 410,000 physician visits, 205,000-272,000 emergency department visits, 55,000-70,000 hospitalizations, and 20 deaths annually, with total annual direct and indirect costs of about $1 billion. Several candidate rotavirus vaccines have been assessed in field trials since 1982, which were developed from a variety of rotavirus strains. In 1998, ACIP recommended Rotashield (RRV-TV), a rhesus-based tetravalent rotavirus vaccine, for routine vaccination of U.S. infants. However, RRV-TV was withdrawn from the U.S. market within 1 year of its introduction because of its association with intussusception. In February 2006, the U.S. Food and Drug Administration (FDA) licensed a live, oral, human-bovine reassortant pentavalent rotavirus vaccine (RotaTeq™) as a 3-dose series for use among infants. The vaccine contains five reassortant rotaviruses developed from human and bovine parent rotavirus strains (serotypes G1, G2, G3, G4, and P1[8]).

Formulating questions and choosing outcomes

The main question is whether pentavalent rotavirus vaccine (RotaTeq™) should be administered routinely to all infants. There are also questions as to whether breastfed infants and infants born prematurely can receive the vaccine. There is also a question as to whether the vaccine can be co-administered with DTaP, Hib, IPV, hepatitis B, and PCV vaccines.

What is the key question to be answered?

Brainstorm

Should rotavirus vaccine be administered routinely to infants?

Population (P): Infants

Intervention (I): Rotavirus vaccine (RotaTeq) administered orally at ages 2, 4, and 6 months.

Control (C): No rotavirus vaccination.

Outcomes (O): Rotavirus diarrhea; severe rotavirus diarrhea; deaths; hospitalizations; office visits; severe adverse events, including intussusception; mild adverse events.

Question:

Should pentavalent rotavirus vaccine be administered routinely to infants at ages 2, 4, and 6 months?

What are the most important outcomes?

Choose the most important outcomes for decision making.
Consider:

- *outcomes that might be important to someone making a decision to use or not to use the interventions (make sure to include both benefits and harms)*
- *outcomes that have been reported in systematic reviews and individual studies*

Rate the relative importance for each outcome on a 9 point scale ranging from 1 (not important) to 9 (critical), regardless of whether data are available for the outcome.

1 – 3 not important and not included in the Evidence Tables
4 – 6 important but not critical for making a decision (inclusion in the Evidence Tables may depend on how many other important outcomes there are, as a total of up to seven important and critical outcomes are generally included in evidence tables)
7 – 9 critical for making a decision and should definitely be included in the Evidence Tables.

The same rating can be used several times (i.e. same number for more than one outcome).
If a surrogate outcome is used, the importance of the corresponding health outcome should be scored.

Outcome	Importance	Include in Evidence Profile table?	
1. Rotavirus diarrhea (any severity)	**6**	Yes	No
2. Severe rotavirus diarrhea	**9**	Yes	No
3. Hospitalizations for rotavirus diarrhea	7	Yes	No
4. Office visits for rotavirus diarrhea	3	Yes	No
5. Deaths from rotavirus diarrhea	2	Yes	No
6. Intussusception	**9**	Yes	No
7. Other serious adverse events (fatal, life threatening, or require hospitalization)	**9**	Yes	No
8. Mild adverse events	3	Yes	No
9. Cost-effectiveness	6	Yes	No

Evidence Retrieval, Assessment and Synthesis

Included Studies: We included three phase 3 clinical trials of the efficacy, immunogenicity, and safety of the pentavalent rotavirus vaccine – Study 006 rotavirus efficacy and safety trial (REST); Study 007 the end-expiry dose trial (end-expiry); and Study 009 immunogenicity study to demonstrate lot consistency of RotaTeq (lot-consistency).

Publications from Study 006 (REST):

- Vesikari T et al. Safety and efficacy of a pentavalent human-bovine (WC3) reassortment rotavirus vaccine. N Engl J Med 2006;354:23-33.
- Goveia MG et al. Safety and efficacy of the pentavalent human-bovine (WC3) reassortant rotavirus vaccine in healthy premature infants. Pediatr Infect Dis J 2007;26(12):1099-1104.
- Goveia MG et al. Efficacy of pentavalent human-bovine (WC3) reassortant rotavirus vaccine based on breastfeeding frequency. Pediatr Infect Dis J 2008;27(7):656-658.
- Rodriguez ZM et al. Concomitant use of an oral live pentavalent human-bovine reassortant rotavirus vaccine with licensed parenteral pediatric vaccines in the United States. Pediatr Infect Dis J 2007;26(3):221-227.
- Clinical Review of New Biologics License Application STN #125122 RotaTeq. FDA, 2006. http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm094063.htm

Publications from Study 007 (end-expiry):

- Block SL et al. Efficacy, immunogenicity, and safety of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine at the end of shelf life. Pediatrics 2007;119(1):11-18.
- Clinical Review of New Biologics License Application STN #125122 RotaTeq. FDA, 2006.

Publications from Study 009 (lot-consistency):

- Clinical Review of New Biologics License Application STN #125122 RotaTeq. FDA, 2006.

Excluded Studies: We excluded the following studies from the evidence review:

- Phase 1 and Phase 2 clinical studies (001, 002, 003, 004, 005) that utilized a different vaccine formulation (fewer serotypes, not buffered, required pre-feed) than what was used in the phase 3 studies 006, 007, 009.
- Trials of the tetravalent human-rhesus RotaShield vaccine (Wyeth Ayerst) that was withdrawn from the U.S. market after licensure in 1998 because of an association with intussusception.
- Phase 1, Phase 2, and Phase 3 trials of a monovalent rotavirus vaccine containing a human rotavirus strain (type G1P1[8]). The monovalent rotavirus vaccine (Rotarix™), produced by GlaxoSmithKline, was subsequently licensed in 2008.
- Trials of candidate rotavirus vaccines using other rotavirus strains (e.g. bovine strain, simian strain, lamb strain).

Table. Characteristics of Included Studies

Author, Year	Methods	Participants	Intervention	Main outcomes	Funding source	Notes
Vesikari, 2006 (Study 006)	RCT. Follow-up for 1 full rotavirus season after vaccination (some followed through a 2nd full RV season)	Infants ages 6–12 weeks. 11 countries, including USA and Finland	3-doses of vaccine compared to placebo	Rotavirus diarrhea, severe rotavirus diarrhea, hospitalization, immunogenicity, serious adverse events, adverse events	Merck	
Block, 2007 (Study 007)	RCT. Follow-up for 1 full rotavirus season after vaccination	Infants ages 6–12 weeks. USA and Finland	3-doses of vaccine compared to placebo	Rotavirus diarrhea, severe rotavirus diarrhea, immunogenicity, serious adverse events, adverse events	Merck	
Anony-mous, 2005 (Study 009)	RCT. Follow-up for 42 days after vaccination for safety	Infants ages 6–12 weeks. USA	3-doses of vaccine compared to placebo	Immunogenicity, serious adverse events, adverse events	Merck	

RCT, Randomized Controlled Trial

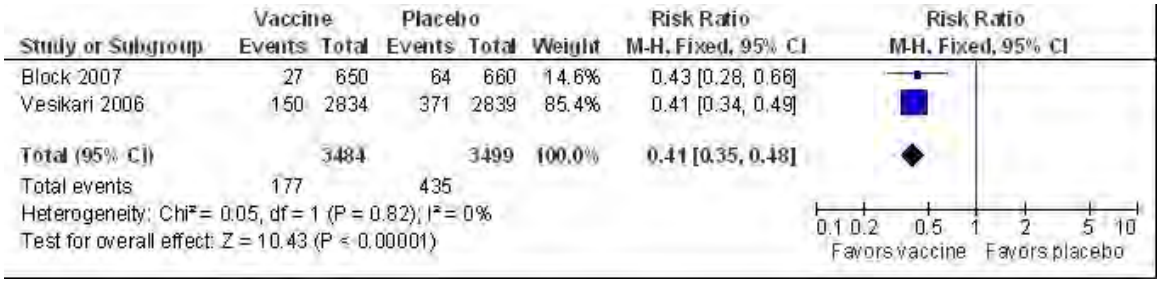
Forest Plots using Data Abstracted from Included Studies
(RevMan software used for analysis).

Question: Routine Administration of Pentavalent Rotavirus Vaccine

Analysis 01.01. Comparison 01 Routine administration, Outcome 01 *Rotavirus diarrhea* occurring through the first full rotavirus season after vaccination

Per-protocol analysis (rotavirus diarrhea occurring ≥14 days after the 3rd dose)

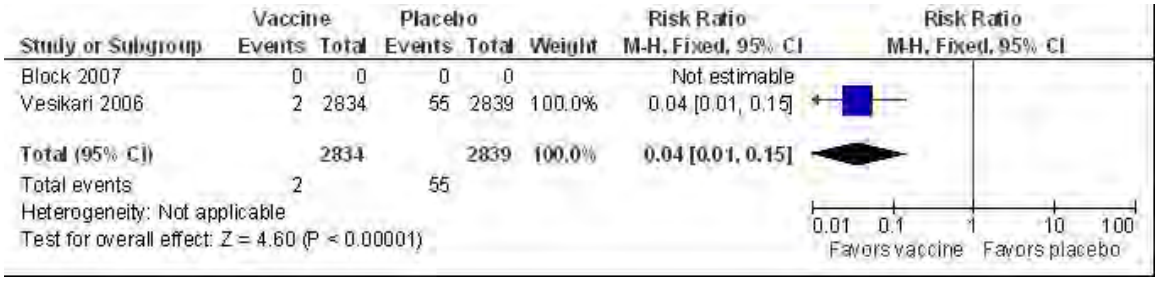
Intention-to-Treat analysis (rotavirus diarrhea occurring anytime after the 1st dose; includes infants who received 1 or 2 doses only)



Analysis 01.02. Comparison 01 Routine administration, Outcome 02 *Severe rotavirus diarrhea* occurring through the first full rotavirus season after vaccination

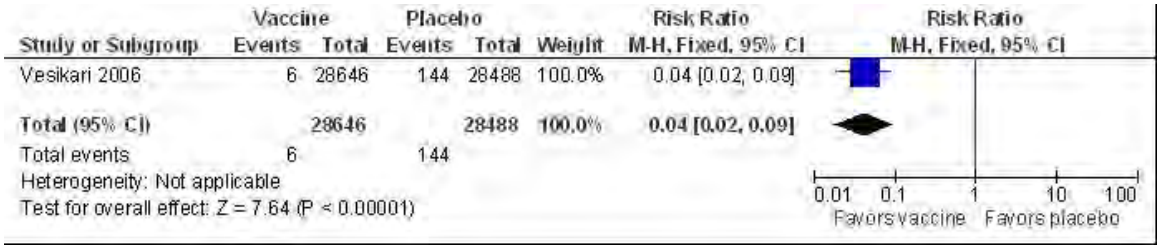
Per-protocol analysis

Intention-to-Treat analysis

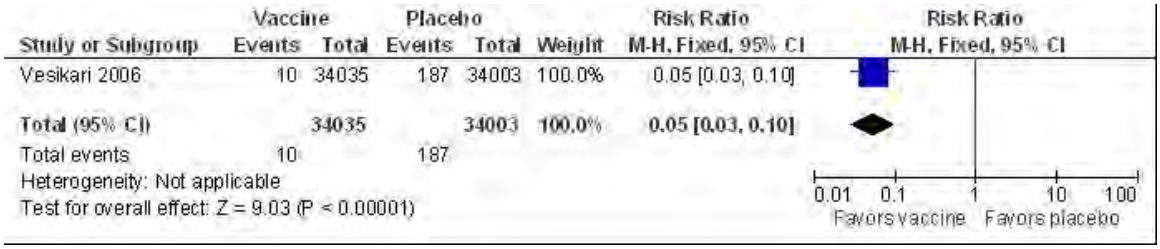


Analysis 01.03. Comparison 01 Routine administration, Outcome 03 *Hospitalization for rotavirus diarrhea* through the first full rotavirus season after vaccination

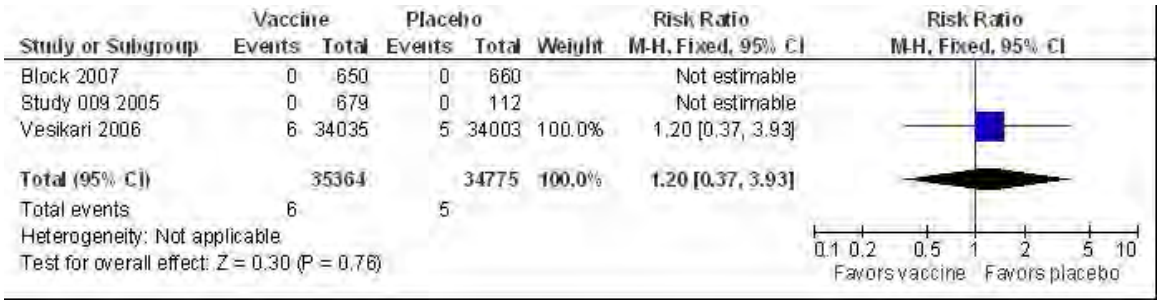
Per-protocol analysis



Intention-to-Treat analysis



Analysis 01.04. Comparison 01 Routine administration, Outcome 05 *Intussusception* within 42 days after any dose



Analysis 01.05. Comparison 01 Routine administration, Outcome 06 *Serious adverse events* (other than intussusception) within 42 days after any dose

Evidence Tables

Table 1. Benefits: Pentavalent Rotavirus Vaccine[a]

Outcome	No. of subjects (# studies)	Incidence in controls	Incidence in vaccinated	Vaccine efficacy (95% CI)	Absolute risk per 1000 (95% CI)	Number Needed to Treat (Vaccinate)
Rotavirus diarrhea	5,627 (2 RCTs)	12.9%	3.5%	73% (66, 78)	-94 (-85, -100)	11
Severe RV diarrhea	5,627 (2 RCTs)	2.0%	0.1%	97% (86, 99)	-19 (-17, -20)	52
Hospitaliza -tion for RV diarrhea	57,134 (1 RCT)	0.5%	0.02%	96% (91, 98)	-5 (-5, -5)	205

RCT, Randomized Controlled Trial. RV, Rotavirus.
[a]Per-protocol analysis.

Table 2. Safety: Pentavalent Rotavirus Vaccine

Outcome	No. of subjects (# studies)	Incidence in controls	Incidence in vaccinated	Relative Risk (95% CI)	Risk Difference per 1000 (95% CI)	Number Needed to Treat[a]
Intussus-ception	70,139 (3 RCTs)	1.4 per 10,000	1.7 per 10,000	1.20 (0.37–3.93)	0.03 (-0.1, 0.4)	-
Other serious adverse events	70,139 (3 RCTs)	2.3%	2.2%	0.96 (0.87–1.06)	-1 (-3, 1)	-

[a]Not meaningful when risk difference is not statistically significant.

Table 3. Type of Evidence: Pentavalent Rotavirus Vaccine

Outcome	Design (# studies)	Risk of bias	Inconsis-tency	Indirect-ness	Impreci-sion	Other considera-tions[a]	Evidence type[b]
Rotavirus diarrhea (RV)	RCT (2)	No serious	No serious	No serious	No serious	None	1
Severe RV diarrhea	RCT (2)	No serious	No serious	No serious	No serious	None	1
Hospitalization for RV diarrhea	RCT (1)	No serious	No serious	No serious	No serious	None	1
Intussuscep-tion	RCT (3)	No serious	No serious	No serious	No serious	None	1
Other serious adverse events	RCT (3)	No serious	No serious	No serious	No serious	None	1

[a]Strength of association, dose-response, opposing plausible residual confounding or bias, publication bias.
[b]Evidence type:
1= Randomized controlled trials (RCTs), or overwhelming evidence from observational studies.
2= RCTs with important limitations, or exceptionally strong evidence from observational studies.
3= Observational studies, or RCTs with notable limitations.
4= Clinical experience and observations, observational studies with important limitations, or RCTs with several major limitations.

Table 4. Summary of Evidence: Pentavalent Rotavirus Vaccine

Comparison	Outcome	Study design (# studies)	Findings	Evidence type	Overall evidence type
Rotavirus vaccination vs. No vaccination	Rotavirus diarrhea (RV)	RCT (2)	Decreased risk among vaccinated	1	1
	Severe RV diarrhea[a]	RCT (2)	Decreased risk among vaccinated	1	
	Hospitalization for RV diarrhea[a]	RCT (1)	Decreased risk among vaccinated	1	
	Intussusception[a]	RCT (3)	No difference	1	
	Other serious adverse events[a]	RCT (3)	No difference	1	

RCT: Randomized controlled trial.

[a] Critical outcome (overall evidence type is based on the critical outcomes).

Moving from Evidence to Recommendation

⇨ **Overall evidence type**

Overall evidence type across all *critical* outcomes	1

⇨ **Values and preferences (assume a set of values for each outcome considered)**

OUTCOME	**VALUES AND PREFERENCES**
Rotavirus diarrhea	Relatively lower value
Severe rotavirus diarrhea	High value
Hospitalization for rotavirus diarrhea	High value
Intussusception	High value
Other serious adverse events	High value
Cost effectiveness	Relatively lower value

⇨ **Draft recommendation**

We recommend vaccination of infants with three doses of rotavirus vaccine.

⇨ Judgments about the recommendation category (Category A; Category B)

Use the table below to make a judgment. The four factors in this table will determine whether the recommendation is category A (recommendation for; recommendation against) or category B (individual clinical decision making). Frequent 'yes' answers increase the likelihood of a category B recommendation.

Factors that can weaken a recommendation	Decision	Explanation
Lower evidence level [The lower the evidence level, the more likely is a category B recommendation.]	☐ **Yes** ☐ **No**	Evidence level is high
Lower net benefit or uncertainty about the balance of benefits versus harms and burdens [The smaller the net benefit and the lower certainty for that benefit, the more likely is a category B recommendation.]	☐ **Yes** ☐ **No**	Benefits are large compared to potential harms
Variability or uncertainty in values [The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely is a category B recommendation.]	☐ **Yes** ☐ **No**	Patients and providers would accept rotavirus vaccination
Lower cost-effectiveness or uncertainty about whether the net benefits are worth the costs [The higher the costs of an intervention – that is, the more resources consumed – the more likely is a category B recommendation]	☐ **Yes** ☐ **No**	The price per dose is not known. Vaccine is likely to be cost-saving from the societal perspective at a total cost of up to $156 per child ($42/dose).[a]

[a]The manufacturer subsequently set the price at $62.50/dose (in 2006 dollars); a rotavirus vaccination program would cost an estimated $197,190 per life-year saved.

If consensus is not reached by discussion, the workgroup can use the table below to record their views (votes) about the recommendation related to a specific intervention, based on their analysis of the available evidence, the benefits and harms, values and preferences and economic analyses. This assessment is then mapped to the recommendation for the use, or non-use, of each intervention.

Insert the number of votes for the recommendation in each category

	Recommend for (Category A)	**Individual decision making (Category B)**	**Individual decision making (Category B)**	**Recommend against (Category A)**
Assessors' view of the balance of desirable and undesirable consequences of the intervention	Desirable consequences clearly outweigh undesirable consequences	Desirable consequences probably outweigh undesirable consequences	Undesirable consequences probably outweigh desirable consequences	Undesirable consequences clearly outweigh desirable consequences
Recommendation	We recommend to "do something"	We suggest to "do something"	We suggest to "not do something"	We recommend to "not do something"
Number of votes in workgroup				

Recommendation category	Category A

⇨ **Final recommendation**

We recommend routine vaccination of U.S. infants with three doses of rotavirus vaccine administered orally at ages 2, 4, and 6 months.

⇨ **Remarks**

Nearly every child in the U.S. is infected with rotavirus by age 5 years, and the majority will have gastroenteritis, resulting in approximately 410,000 physician visits, 205,000-272,000 emergency department visits, and 55,000-70,000 hospitalizations each year. Randomized clinical trials show that vaccination reduces severe rotavirus gastroenteritis. Benefits are substantial compared to potential harms.

Appendix 2. Data Extraction Forms

Form 1. Information on included studies

Author, Year:
Name of reviewer: Date completed:

I. Methods
Study design:
Number randomized or enrolled (total and per group):
Number analyzed (total and per group):
Losses to follow up (for each outcome):

II. Participants
Setting:
Country:
Age:
Gender (% female):
Race/ethnicity:
Inclusion criteria:
Exclusion criteria:
Equivalence of baseline characteristics:

III. Interventions
Intervention group:
Comparison group:

IV. Outcomes

V. Notes
Type of study (published/unpublished):
Funding source:
Study period:
Reported subgroup analyses:

Form 2a. Assessment of risk of bias for randomized controlled trials[a]

Author, Year:

Name of reviewer: Date completed:

Criteria	Description	Yes /No/ Unclear	Quote from study
Adequate allocation sequence generation	The investigators describe a random component in the sequence generation process (e.g. computer random number generator). Problem if "pseudo" or "quasi" randomization with allocation by day of week, birth date, chart number, etc.		
Adequate allocation sequence concealment	Those enrolling patients cannot foresee the group to which the next enrolled patient will be allocated (e.g. central allocation, sequentially numbered sealed envelopes)		
Adequate blinding of participants and personnel	Study participants and personnel are not aware of the arm to which patients are allocated		
Adequate blinding of outcome assessors	Outcome assessors are not aware of the arm to which patients are allocated (assess separately for each outcome; outcomes may be grouped as subjective and objective)		
Incomplete outcome data addressed	Loss to follow-up; adherence to the intention to treat principle when indicated[b] (assess separately for each outcome; outcomes may be grouped as short-term and long-term)		
Free of selective outcome reporting	Study reports all pre-specified or expected outcomes. Problem if reporting of some outcomes and not others on the basis of the results		
Free of other biases	For example: extreme baseline imbalance; differential diagnostic activity		

[a]See Cochrane handbook, Chapters 8 and 16 (http://www.cochrane-handbook.org/).
[b]Intention-to-treat analyses may not be appropriate for adverse events or non-inferiority studies (ref: Cochrane handbook, Chapter 16, Section 16.2.1).
For observational studies, see http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm.
For controlled before-after and interrupted time series, see www.epoc.cochrane.org/epoc-author-resources.

Form 2b. Assessment of risk of bias for cluster-randomized trials[a]

Author, Year:

Name of reviewer: Date completed:

Criteria	Description	Yes /No/ Unclear	Quote from study
Recruitment bias	Individuals are recruited to the trial after the clusters have been randomized; knowledge of whether each cluster is an intervention or control cluster could affect the types of participants recruited. Strategies to minimize the possibility of selection bias include inclusion of all individuals within a cluster or recruitment of individuals by a person masked to the cluster allocation.[b]		
Baseline imbalance	Baseline imbalance between the randomized groups, in terms of either the clusters or the individuals. Statistical adjustment for baseline characteristics can help reduce concern about the effects of baseline imbalance.		
Loss of clusters	Loss of clusters from a trial may lead to bias. In addition, missing outcomes for individuals within clusters may lead to a risk of bias. For example, differential adherence and follow-up can occur for whole clusters or individuals in a cluster.		
Incorrect analysis	The clustering effect is not taken into account in the analysis. Such analyses produce erroneously narrower standard errors that will result in too much weight in a meta-analysis if they remain uncorrected.		
Comparability with individually randomized trials	In a meta-analysis including both cluster and individually randomized trials, or including cluster-randomized trials with different types of clusters, possible differences between the intervention effects being estimated [because of herd effect] need to be considered. If intervention effect is still demonstrated in individually randomized trials, a confident conclusion about the presence of an effect can be drawn. Herd effects may be different for different types of cluster.		
Other biases			

[a]See Cochrane handbook, Chapter 16.3 (http://www.cochrane-handbook.org/).

[b]Campbell MK et al. CONSORT Statement: extension to cluster randomised trials. BMJ 2004;328:702-8.

Form 3a. Data abstraction for dichotomous outcomes

Author, Year:
Name of reviewer:
Date completed:

Outcome (as stated in study)				
	Intervention group		Comparison group	
Time at follow-up	n (Number with outcome)	N (Total number in group)	n (Number with outcome)	N (Total number in group)
Notes:				

Outcome (as stated in study)				
	Intervention group		Comparison group	
Time at follow-up	n (Number with outcome)	N (Total number in group)	n (Number with outcome)	N (Total number in group)
Notes:				

Outcome (as stated in study)				
	Intervention group		Comparison group	
Time at follow-up	n (Number with outcome)	N (Total number in group)	n (Number with outcome)	N (Total number in group)
Notes:				

Form 3b. Data abstraction for continuous outcomes

Author, Year:
Name of reviewer:
Date completed:

Outcome (as stated in study)						
	Intervention group			Comparison group		
Time at follow-up	n	Mean (95% CI)	SD[a]	n	Mean (95% CI)	SD[a]
Notes:						

Outcome (as stated in study)						
	Intervention group			Comparison group		
Time at follow-up	n	Mean (95% CI)	SD[a]	n	Mean (95% CI)	SD[a]
Notes:						

Outcome (as stated in study)						
	Intervention group			Comparison group		
Time at follow-up	n	Mean (95% CI)	SD[a]	n	Mean (95% CI)	SD[a]
Notes:						

[a]Standard deviation (SD) = Standard error * square root of total participants in group (n). If SD is not reported, it can be computed from 95% CI (see Cochrane handbook, Chapter 7, Section 7.7.3).

Form 4. Determining evidence type

Name of reviewer:
Date completed:

Criteria	**Assessment** (circle one for each criterion)	**Reasons for assessment**	**Evidence type**[a] (Circle one per outcome)
OUTCOME:			
Risk of bias	No serious (-1) very serious (-2)		1
Inconsistency	No serious (-1) very serious (-2)		2
Indirectness	No serious (-1) very serious (-2)		3
Imprecision	No serious (-1) very serious (-2)		4
Publication bias	Unlikely likely (-1) very likely (-2)		
Strength of association	No Large (+1) Very large (+2)		
Dose response relation	No Yes (+1)		
Opposing plausible residual confounding or bias	No Yes (+1)		

[a]Evidence type:

1= Randomized controlled trials (RCTs), or overwhelming evidence from observational studies.
2= RCTs with important limitations, or exceptionally strong evidence from observational studies.
3= Observational studies, or RCTs with notable limitations.
4= Clinical experience and observations, observational studies with important limitations, or RCTs with several major limitations.

Appendix 3. Using GRADEpro Software for Preparing Evidence Tables

(Available at http://www.gradeworkinggroup.org/toolbox/index.htm)

1. Open GRADEpro.

2. Choose New Profile in the welcome screen, enter file name, and click **Save**.

3. Create profile group (usually overall topic of guidelines or main question)

4. Create a profile (specific health care question)

    a. Choose the format of the question in the drop down box (e.g. Should intervention versus comparison be used for health situation?)
    b. Add information about the comparison, intervention, setting, etc.
    c. Add bibliographic information about the studies or reviews used to create the profile.

5. Create the outcomes by naming them.

6. Select an outcome. For each outcome there are 2 sections: Summary of Findings screen and the Quality Assessment screen.

7. Select the Quality Assessment screen.

    a. Complete it by first confirming number and type of studies
    b. Assess the quality of evidence for the outcome. Downgrade or upgrade evidence according to GRADE criteria and enter footnotes when necessary.

8. Select Summary of Findings screen.
    a. Add data about participants, estimate of effect, baseline risks, etc.

9. Repeat (#7 and #8) for all outcomes in the profile.

10. Preview GRADE evidence profile, double check presentation and edit if necessary.

11. Export the table to a document in word, html, image, etc.

Appendix 4. Using Review Manager (RevMan) Software

(Available at http://ims.cochrane.org/revman)

The RevMan software, developed by the Nordic Cochrane Centre, can be used to generate forest plots, funnel plots, and risk of bias graphs. Forest plots display results of individual studies in a graphical manner, and include pooled estimates of effects (e.g. summary risk ratios across studies) and tests for heterogeneity and I^2. Funnel plots can help assess publication bias.

A tutorial is available on using RevMan. After opening RevMan, click **Help**, select **Tutorial**, and then click **Intervention Reviews**. For generating the plots and graphs, the first step is to add studies as described in Part 3 of the tutorial. Part 5 of the tutorial describes how forest plots and funnel plots can be generated, and Part 4 describes the procedure to create risk of bias graphs.

ACIP Evidence Based Recommendations Work Group Members

Temte, Jonathan, M.D., Ph.D. (Chair)
ACIP Member 2008-present
Associate Professor, Department of Family Medicine
University of Wisconsin School of Medicine and Public Health
Madison, WI

Beck, Robert L., J.D.
ACIP Member 2005-2009
Lake Monticello
Palmyra, VA

Calonge, Ned, M.D., M.P.H.
Chief Medical Officer, Colorado Department of Public Health and Environment
Associate Professor, University of Colorado Health Sciences Center
Denver, CO

Campos-Outcalt, Doug, M.D., M.P.A.
ACIP Member 2011-present
Associate Head, Department of Family & Community Medicine
University of Arizona College of Medicine
Phoenix, AZ

Huang, Jiangcheng, M.D., M.S.
Director, Immunization Program
Maine Department of Health and Human Services
Augusta, ME

Kinsinger, Linda, M.D., M.P.H.
ACIP *Ex Officio* Member, Department of Veterans Affairs
Director, National Center for Health Promotion and Disease Prevention
Patient Care Services
Veterans Health Administration
Durham, NC

Langley, Joanne, M.D.
ACIP Liaison Representative, National Advisory Committee on Immunization, Canada
Professor of Pediatrics, IWK Health Center, Infectious Diseases
Halifax, Canada

Marcuse, Edgar M.D., M.P.H.
ACIP Member 2004-2006
Professor Pediatrics, Adjunct Professor Epidemiology
University of Washington
Seattle Children's Hospital
Seattle, WA

Moyer, Virginia M.D., M.P.H.
Professor, Department of Pediatrics
Baylor College of Medicine
Houston, TX

Qaseem, Amir, M.D., Ph.D, M.H.A.
American College of Physicians
Philadelphia, PA

Schaffner, William, M.D.
ACIP Liaison Representative, National Foundation for Infectious Diseases
Professor and Chair Department of Preventive Medicine
Vanderbilt University School of Medicine
Nashville, TN

Schunemann, Holger, M.D., Ph.D., M.Sc., FRCP(C)
Professor and Chair, Department of Clinical Epidemiology and Biostatistics
Michael Gent Chair in Healthcare Research
McMaster University Health Sciences Centre
Hamilton, Canada

Tan, Litjen (L.J), Ph.D., M.S.
ACIP Liaison Representative, American Medical Association
Director, Infectious Diseases, Immunology, & Molecular Medicine
American Medical Association
Chicago, IL

Gidudu, Jane, M.D., M.P.H.
Immunization Safety Office
Centers for Disease Control and Prevention
Atlanta, GA

Janes, Gail, Ph.D., M.S.
Division of HIV/AIDS Prevention
Centers for Disease Control and Prevention
Atlanta, GA

Smith, Jean Clare, M.D., M.P.H.
National Center for Immunization and Respiratory Diseases
Centers for Disease Control and Prevention
Atlanta, GA

Ahmed, Faruque, Ph.D.
Immunization Services Division
National Center for Immunization and Respiratory Diseases
Centers for Disease Control and Prevention
Atlanta, GA