

D4.4

Generic Statistical Analysis Plan: combining information on Influenza Vaccine Effectiveness across study sites

777363 - DRIVE

Development of robust and
innovative vaccine
effectiveness

**WP4 – Framework for
analysis and study reports**

Lead contributor	Kaatje Bollaerts (P95) kaatje.bollaerts@p-95.com
Other contributors	Jos Nauta (Abbott), Margarita Riera (P95)

Due date	31 March 2018
Delivery date	19 April 2018
Deliverable type	R
Dissemination level	PU

Document History

Version	Date	Description
	07 11 2017	Outline - first draft
V1.0	29 01 2018	First draft
	22 02 2018	Comments Marga Riera and Jos Nauta
V1.1	26 02 2018	Second draft
	09 03 2018	Comments WP4 team members
V1.2	12 03 2018	First version, to share with SC
	26.03.2018	Comments SC: Seqirus, SP, GSK, IRD, THL
V1.3	29.03.2018	Final version
	18.04.2018	Addition of paragraph about sample size considerations and requirements following discussion with SC

TABLE OF CONTENTS

Document History	2
TABLE OF CONTENTS	3
LIST OF ABBREVIATIONS	4
1 BACKGROUND	5
2 REFERENCE DOCUMENTS.....	6
3 AGGREGATED DATA META-ANALYSIS.....	6
3.1 Objective(s).....	6
3.2 Effect measures.....	6
3.3 Sample size considerations.....	6
3.4 Strategy for data synthesis.....	6
3.4.1 Inclusion criteria.....	6
3.4.2 Meta-analysis.....	7
3.4.3 Outlier and influence analysis	7
3.4.4 Quantifying between-study heterogeneity.....	7
3.4.5 Exploring sources of heterogeneity	7
3.4.6 Sensitivity analysis.....	8
3.4.7 Presentation of results	8
4 REFERENCES.....	10

LIST OF ABBREVIATIONS

AD-MA	Aggregated data meta-analysis
CI	Confidence interval
DRIVE	Development of Robust and Innovative Vaccine Effectiveness
IPD-MA	individual participant data meta-analysis
IVE	Influenza vaccine effectiveness
OR	Odds ratio
RR	Relative risk
SAP	Statistical Analysis Plan
RRR	Relative Risk Ratio

1 BACKGROUND

The DRIVE consortium aims to enable the collaboration of different public and private stakeholders to perform annual brand-specific influenza vaccine effectiveness (IVE) studies for various influenza vaccines on the European market. To this end, IVE studies will be conducted at various study sites across Europe. In a second step, the site-specific data will be combined to obtain overall estimates at the European level. The purpose of this document is to provide guidance for writing the Statistical Analysis Plan (SAP) of combining and presenting information on IVE from different study sites. This document will be updated following the learnings from the pilot year 2017-2018.

There are two statistical approaches for pooling data: a one-stage or a two-stage pooling approach (1). The two-stage approach refers to the classical meta-analytical approach, also called aggregated data meta-analysis (AD-MA). In this approach, the patient-level or minimally aggregated data from each study are analysed separately in order to obtain the effect estimates of interest (here vaccine effectiveness estimates) and the corresponding confidence intervals (CIs). Then, in the second step, the effect estimates are combined by an appropriate meta-analysis model to obtain the meta-analytical (weighted averaged) estimate. The one-stage pooling approach analyses all the combined patient-level or minimally aggregated data from the different data sources in a single step. This approach is also called the individual participant data meta-analysis (IPD-MA).

We opt to pool data using the AD-MA approach, given the statistical equivalence of AD-MA and IPD-MA, given that many of the mentioned advantages of IPD-MA (i.e. transforming data to common sources or measures and standardizing analysis) can also be achieved through harmonization/standardization of the individual site-specific studies and given the additional complexity of performing IPD-MA when data are collected using different study designs (1). Within AD-MA, we prefer the use of random effects meta-analysis model, which assumes that the observed effect estimates can vary across study sites because of differences in the treatment effect in each study site (e.g. due to differences in population, in health care utilization, in circulating influenza strains) as well as sampling variability.

This document builds further upon or relates to the DRIVE generic study protocols for the analyses and presentation of data collected at a single study site, the DRIVE data management plan and the DRIVE report template (see Reference documents).

2 REFERENCE DOCUMENTS

[Here: refer to the generic study protocols for the analyses of the study site-specific data, the data management plan and the report template]

3 AGGREGATED DATA META-ANALYSIS

3.1 Objective(s)

To estimate seasonal IVE (%) through pooling site-specific estimates obtained as described in the site-specific protocols.

[Describe the primary and secondary objectives as per study protocol mentioned in Section 2 and for which pooling will be performed]

3.2 Effect measures

The effect measures for pooling are the study site-specific IVE estimates and their 95% confidence intervals (CIs).

3.3 Sample size considerations

[Sample size considerations for the primary objective(s) should be discussed in this section including the assumptions made for vaccination coverage, vaccine effectiveness and influenza attack rate. This section will be updated pending consultation with the DRIVE Ethics Advisory Board and EMA on the need to establish minimum sample size and/or minimum precision for the primary objective(s)].

3.4 Strategy for data synthesis

3.4.1 Inclusion criteria

We will pool seasonal IVE estimates from the individual study sites in line with the objectives as per study protocol (Section 3.1). Estimates that are not obtained following the study protocols will not be retained for the primary meta-analysis, but might be considered for inclusion as part of a sensitivity analysis (Section 3.4.6). Whenever there are two or more site-specific estimates retained, a meta-analysis will be performed.

Further pooling (e.g. incorporating IVE estimates which were not minimally adjusted for confounding as per study protocol) might be considered upon lack of heterogeneity (see Sections 3.4.4 and 3.4.5).

3.4.2 Meta-analysis

For every objective listed in Section 3.1, a meta-analysis will be performed. First, the study site-specific IVE estimates will be back-transformed to the original relative risk (RR) estimates (in case of cohort studies) and odds ratio (OR) estimates (in case of case-control studies), which will be subsequently log-transformed, or

$$\log RR \text{ or } \log OR = \log(1 - VE)$$

Then, standard inverse variance weighted random-effects meta-analysis of the log-transformed RR and OR estimates will be used to obtain the pooled estimate (2). The pooled estimate (and 95% CI) will then be back-transformed to obtain the pooled IVE estimate (and 95% CI), expressed in %.

3.4.3 Outlier and influence analysis

For every meta-analysis performed, the potential impact of outliers and influential estimates on the pooled estimate will be evaluated. Studentized deleted residuals r will be used to identify outliers in the meta-analysis. Site-specific IVE estimates will be considered outlying from meta-analysis when $|r| > 2.5$, where $|r|$ indicates the absolute value of the residual (3).

The standardized DFBETAs statistic will be used to identify influential estimates, examining the change in the averaged IVE from the random-effects model when excluding one site-specific estimate in turn. Site-specific estimates will be considered influential from meta-analysis when $|DFBETAs| > 2/\sqrt{n}$, where $|DFBETAs|$ indicates the absolute value of the DFBETAs statistics and n is the number of effect estimates (3).

Site-specific estimates that are outlying and influential, will be excluded from meta-analysis and the reason for being outlying will be investigated and documented.

3.4.4 Quantifying between-study heterogeneity

An indication for the heterogeneity among estimates from different study sites will be obtained by calculating I^2 according to Higgins et al (4). The I^2 statistic is to be interpreted as the proportion of total variation in the estimates of treatment effect that is due to heterogeneity between studies. Low, moderate and high levels of heterogeneity correspond to I^2 values of 25%, 50% and 75% respectively. In case I^2 is high, it is worthwhile to explore sources of heterogeneity (Section 3.4.5).

3.4.5 Exploring sources of heterogeneity

In case of at least 5 site-specific IVE estimates, stratified analyses and meta-regression might be used to explore whether the magnitude of the IVE estimates are associated with design or other

characteristics of the study site-specific estimates of interest (e.g. study design, adjustments for certain covariates). In stratified analyses, the meta-analysis (as in Section 3.3.3) will be repeated for each stratum of characteristics separately. In meta-regression, the meta-analysis (as in Section 3.3.3) will be extended with the site-specific study characteristics as predictor variables and relative risk ratios (RRRs) will be obtained (5). For example, assume the characteristic of interest is study design (cohort vs case-control studies). Then, the RRRs is to be interpreted as the ratio of the pooled IVE estimate of the case-control studies to the pooled IVE estimate of the cohort studies.

The permutation test as proposed by Higgins et al (6) will be used to assess the significance of a study characteristic while controlling the risk of false-positive results. If the study characteristic is not statistically significant in the meta-regression model, the study characteristic is unlikely a source of heterogeneity, and pooling across that study characteristic might be considered.

3.4.6 Sensitivity analysis

Sensitivity analysis in line with the study protocol will be performed.

Additional sensitivity analyses will be performed by including site-specific estimates that were excluded from the main meta-analysis models because 1) they were not obtained following the study-protocol (Section 3.3.1) or 2) they were identified as outlying and influential (Section 3.3.3).

3.4.7 Presentation of results

The site-specific IVE estimates (and 95% CIs) will be presented using a forest plot complemented with the pooled IVE estimate (and 95% CIs) as outlined in the report template. Estimates that were excluded from meta-analysis will included in the forest plot, but these estimates will be tagged as excluded. An example of a forest plot with pooled estimates by setting is given in Figure 1. This plot is generated using artificial data based on cohort designs.

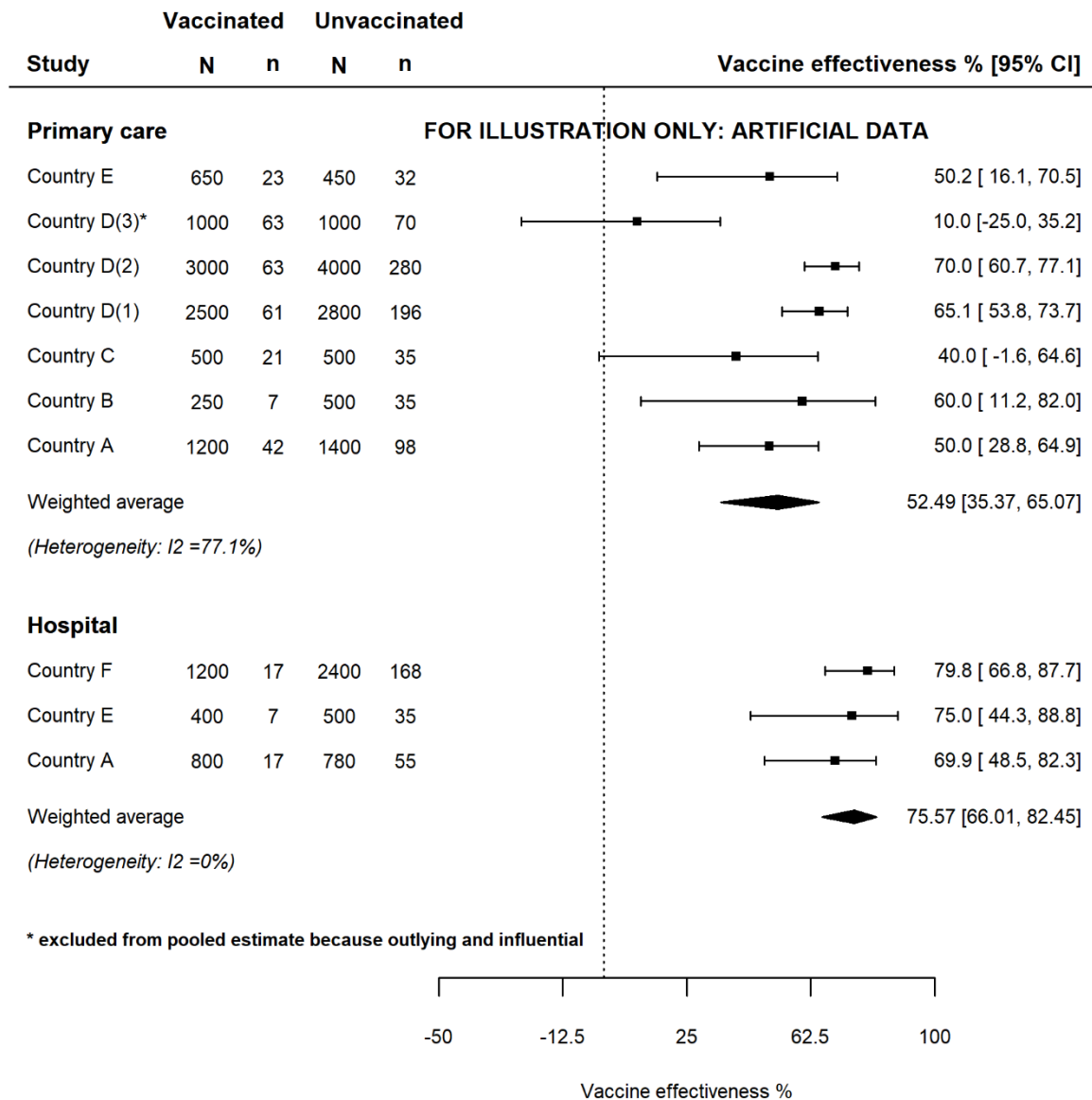


Figure 1: Forest plot and meta-analyses of influenza vaccine effectiveness, by health care setting. This plot is generated using artificial data based on cohort designs

4 REFERENCES

1. Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med* 2017;36(5):855-875.
2. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-88.
3. Viechtbauer W, Cheung MW. Outlier and influence diagnostics for meta-analysis. *Res Synth Methods* 2010;1(2):112-25.
4. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21(11):1539-58.
5. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Methods* 2010;1(2):97-111.
6. Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004;23(11):1663-82.