Taylor & Francis
Taylor & Francis Group

Expert Review of Vaccines

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ierv20

Mapping and ranking outcomes for the evaluation of seasonal influenza vaccine efficacy and effectiveness: a delphi study

Chiara de Waure, Elisabetta Alti, Vincenzo Baldo, Paolo Bonanni, Michele Conversano, Alberto Fedele, Giovanni Gabutti, Roberto Ieraci, Francesco Landi, Raffaele Landolfi, Andrea Orsi, Caterina Rizzo, Alessandro Rossi, Alberto Villani, Francesco Vitale & Alexander Domnich

To cite this article: Chiara de Waure, Elisabetta Alti, Vincenzo Baldo, Paolo Bonanni, Michele Conversano, Alberto Fedele, Giovanni Gabutti, Roberto Ieraci, Francesco Landi, Raffaele Landolfi, Andrea Orsi, Caterina Rizzo, Alessandro Rossi, Alberto Villani, Francesco Vitale & Alexander Domnich (2024) Mapping and ranking outcomes for the evaluation of seasonal influenza vaccine efficacy and effectiveness: a delphi study, Expert Review of Vaccines, 23:1, 636-644, DOI: 10.1080/14760584.2024.2367457

To link to this article: https://doi.org/10.1080/14760584.2024.2367457

EXPERT REVIEW OF VACCINES
2024, VOL. 23, NO. 1, 636–644
https://doi.org/10.1080/14760584.2024.2367457

Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH

OPEN ACCESS Check for updates

Mapping and ranking outcomes for the evaluation of seasonal influenza vaccine efficacy and effectiveness: a delphi study

Chiara de Waure [a], Elisabetta Alti[b], Vincenzo Baldo [c], Paolo Bonanni [d], Michele Conversano [e], Alberto Fedele [f], Giovanni Gabutti [g], Roberto Ieraci [h], Francesco Landi [i], Raffaele Landolfi[j], Andrea Orsi [k,l], Caterina Rizzo [m], Alessandro Rossi[n], Alberto Villani [o,p], Francesco Vitale [q] and Alexander Domnich [l]

[a]Department of Medicine and Surgery, University of Perugia, Perugia, Italy; [b]Department of General Practice, Local Health Unit Toscana Centro, Florence, Italy; [c]Department of Cardiac, Thoracic, Vascular Sciences and Public Health, University of Padua, Padua, Italy; [d]Department of Health Sciences, University of Florence, Florence, Italy; [e]Prevention Department, Local Health Authority of Taranto, Taranto, Italy; [f]Prevention Department, Hygiene and Public Health Service, Local Health Unit, Lecce, Italy; [g]Coordinator Working Group "Vaccines and Immunization Policies", Italian Scientific Society of Hygiene, Preventive Medicine and Public Health (SItI), Cogorno (Ge), Italy; [h]Vaccination strategies group, Lazio region, Italy; [i]Department of Geriatrics and Orthopedics, Fondazione Policlinico Universitario "A. Gemelli", IRCCS, Università Cattolica del Sacro Cuore, Rome, Italy; [j]Department of Internal Medicine, Università Cattolica del Sacro Cuore, Rome, Italy; [k]Department of Health Sciences, University of Genoa, Genoa, Italy; [l]Hygiene Unit, San Martino Policlinico Hospital – IRCCS for Oncology and Neurosciences, Genoa, Italy; [m]Department of Translational Research and New Technologies in Medicine and Surgery, University of Pisa, Pisa, Italy; [n]Italian College of General Practitioners and Primary Care, Florence, Italy; [o]Pediatric Unit, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; [p]System Medicine Department, Tor Vergata University of Rome, Rome, Italy; [q]Department of Health Promotion, Maternal and Infant Care, Internal Medicine and Medical Specialties (PROMISE), University of Palermo, Palermo, Italy

ABSTRACT

Background: Protection provided by seasonal influenza vaccination (SIV) may be measured against numerous outcomes, and their heterogeneity may hamper decision-making. The aim of this study was to explore outcomes used for estimation of SIV efficacy/effectiveness (VE) and obtain expert consensus on their importance.

Research design and methods: An umbrella review was first conducted to collect and map outcomes considered in systematic reviews of SIV VE. A Delphi study was then performed to reach expert convergence on the importance of single outcomes, measured on a 9-point Likert scale, in principal target groups, namely children, working-age adults, older adults, subjects with co-morbidities and pregnant women.

Results: The literature review identified 489 outcomes. Following data reduction, 20 outcomes were selected for the Delphi process. After two Delphi rounds and a final consensus meeting, convergence was reached. All 20 outcomes were judged to be important or critically important. More severe outcomes, such as influenza-related hospital encounters and mortality with or without laboratory confirmation, were generally top-ranked across all target groups (median scores ≥8 out of 9).

Conclusions: Rather than focusing on laboratory-confirmed infection per se, experimental and observational VE studies should include more severe influenza-related outcomes because they are expected to exercise a greater impact on decision-making.

ARTICLE HISTORY
Received 14 December 2023
Accepted 10 June 2024

KEYWORDS
Influenza; influenza vaccines; vaccine efficacy; vaccine effectiveness; outcome assessment; Delphi method

1. Introduction

Seasonal influenza vaccination (SIV) is a major public health intervention able to reduce the burden of disease [1]. SIV is a cost-effective intervention even if vaccine efficacy/effectiveness (VE) is as low as 4% [2]. Indeed, SIV VE is still suboptimal [3], especially in older adults and against the A(H3N2) subtype [4] and depends on a variety of factors related to virus, vaccine, and vaccinee [5].

SIV VE may be established against numerous outcome measures, the choice of which may have a profound effect on VE estimates [6]. Traditionally, laboratory-confirmed influenza (LCI) is considered a gold standard outcome for the assessment of VE in both outpatient and inpatient settings [3]. The main advantage of LCI is its high specificity. On the other hand, the use of LCI in VE studies is associated with some notable limitations. First, patient enrollment usually occurs soon after the onset of symptoms, while those who had developed an influenza-related complication (e.g. secondary bacterial pneumonia) may be either systematically excluded [3] or test negative due to the natural viral clearance [7]. Second, influenza virus may be responsible for a number of extra-respiratory complications, including its ability to trigger major cardiovascular events [8], for which influenza is typically not considered in the diagnostic routine [7]. Third, routine microbiological testing is not recommended for all subjects seeking care for influenza-like illness (ILI) [9,10] and therefore VE studies based on administrative registers may be underpowered or affected by bias (e.g. physicians may

CONTACT Chiara de Waure chiara.dewaure@unipg.it Department of Medicine and Surgery, University of Perugia, Piazza Lucio Severi 1 6132, Perugia, Italy
Supplemental data for this article can be accessed online at https://doi.org/10.1080/14760584.2024.2367457.

preferentially prescribe laboratory testing to unvaccinated patients or those presenting a more typical clinical picture [3]). Taken together, these empirical concepts point out that under several circumstances, sensitivity of LCI may be low. Indeed, choosing appropriate VE outcomes requires a trade-off between identifying less biased ones and providing information that might be relevant for decision-makers [3].

Simulation studies [11,12] suggest that misclassification of influenza due to imperfect test sensitivity and specificity leads to biased absolute VE (i.e. vaccinated versus unvaccinated) estimates in all principal study designs. This is particularly true for the specificity: the lower the specificity of an influenza-related outcome is, the lower the absolute VE would be [12]. It has been estimated that if the influenza attack rate is 5%, the true VE against LCI is 50% and the risk ratio of pneumonia following LCI is 5.0, the resulting VE against pneumonia would be 8% only [13]. Much less is instead known about the effect of sensitivity and specificity on the relative VE, which is the extent to which one vaccine is more or less effective than another. Indeed, the so-called 'enhanced' quadrivalent formulations, such as MF59-adjuvanted, high-dose, and recombinant vaccines, are currently available and may provide additional benefits over standard-dose non-adjuvanted vaccines in older adults [14]. A simulation study by Lewis et al. [15] has come to an important insight: the magnitude of the established relative VE against less specific outcomes may dramatically change data interpretability. For instance, when the observed relative VE of enhanced versus standard vaccines against all-cause cardiovascular hospitalizations was <5%, the population benefit of enhanced formulations was plausible. By contrast, when the relative VE was set to 18%, unrealistic estimates of additionally averted all-cause cardiovascular hospitalizations were generated [15].

Recent availability of different vaccine types has led to a steady growth in the number of relative VE studies (reviewed in [16–18]). Most relative VE studies were based on retrospective assessment of administrative data claims, for which no laboratory-confirmed outcomes were available. Indeed, a single retrospective cohort study could report relative VE estimates against up to 11 different influenza-related outcomes that may move in opposite directions [18]. This heterogeneity of outcomes and findings may disorient decision-making. Based on this foundation, in this study, we aimed to map outcomes that have been most frequently used for the evaluation of SIV VE and to obtain an expert-driven evaluation of their importance to eventually issue a set of critically important outcomes. This set is meant to be taken in consideration both in the planning and conduction of SIV VE studies on different population groups and in the development of vaccination recommendations at national level.

2. Methods

2.1. Literature review

Within this mixed-method study, the goal of the literature review was only to explore and map outcomes used to evaluate SIV VE in the light of issuing a preliminary list of outcomes for the Delphi study. Considering an increasing number of systematic reviews (SRs) on SIV VE, we conducted an umbrella review of the available SRs [19]. To be included, a record had to be a SR with or without meta-analysis or a protocol for a SR, which aimed to evaluate absolute or relative SIV VE against at least one endpoint. There were no restrictions regarding the study population, setting, design of primary studies, vaccine type, or regimen. SRs focused only on immunogenicity, safety, cost-effectiveness, vaccine acceptance, uptake determinants, and similar were instead excluded. Manuscripts that did not qualify as SRs were also excluded.

Automatic search was performed on 3 March 2023, by AD. For this purpose, PubMed, Scopus, and Cochrane Library (Cochrane Reviews only) were systematically searched using a combination of MeSH (medical subject headings) and title/abstract/keywords free search terms (Table S1). No filters or other restrictions were applied. Following removal of duplicates, titles, and abstracts of the available records were screened by AD and cross-checked by CdW. Separately published protocols and associated SRs were considered as a single record. For review updates or living SRs performed by the same research group, the latest version was considered. The full text of potentially eligible SRs was then assessed independently by AD, and CdW and disagreements were solved by consensus. Finally, a manual search of the included SRs was performed by the backward cross-reference checking.

2.2. Selection and mapping of outcomes

The selected set of SRs underwent the process of data extraction. In particular, all outcomes that were pre-specified in the article methods and/or results were first tabulated. SIV target populations could be one of the following: general population (≥6 months), children (6 months to 14/17 years depending on the definition adopted in the SR), working-age adults (15/18 to 60/64 years depending on the definition adopted in the SR), older adults (≥60/65 years depending on the definition adopted in the SR), pregnant women and/or newborns (through maternal immunization), and subjects of any age with underlying health conditions.

Once extracted, the resulting pool of outcomes underwent a first process of data reduction, i.e. clinically, pathogenetically and semantically similar outcomes were grouped together (e.g. non-fatal myocardial infarction and stroke were combined into a single outcome of 'major cardiovascular events;' different definitions of ILI were combined into a single outcome 'ILI'). Then, on the basis of laboratory confirmation and clinical features, the outcomes were categorized a posteriori into seven mutually exclusive categories, namely: (i) laboratory-confirmed outcomes [e.g. any LCI, severe acute respiratory infection (SARI) with laboratory confirmation, hospital encounter for LCI]; (ii) influenza-related outcomes with no or unknown laboratory confirmation [e.g. ILI, acute respiratory infection (ARI), influenza-related primary care visits, hospital encounters, or deaths derived from administrative data flows); (iii) respiratory complications potentially triggered by influenza virus [e.g. pneumonia and other lower respiratory tract diseases (LRTDs), exacerbation of chronic obstructive pulmonary disease, otitis media]; (iv) cardiovascular complications

potentially triggered by influenza virus (e.g. myocardial infarction, stroke, and other major cardiovascular events); (v) hospital encounters not directly related to influenza (e.g. all-cause hospital encounters); (vi) mortality not directly related to influenza (e.g. all-cause mortality) and (vii) miscellaneous (outcomes that did not fall into the previous categories). Notably, only the first group includes outcomes with explicitly stated laboratory confirmation.

Relative percentages were then calculated within each category considering, as denominator, the total number of SRs assessing any outcome in each of the seven categories. This allowed identifying the most frequently investigated outcomes within each category. This analysis was first performed overall, i.e. considering all the included SRs independently by the study population, and then according to the target population. Indeed, as because several outcomes may be relevant only to some target populations, a subsequent data reduction step was performed according to the target groups described above. Finally, the most common outcomes within each category and target population group were included in the preliminary list for the Delphi process based on the ranking of the calculated relative percentages (from the highest to the lowest).

2.3. Delphi study

A Delphi study was designed to reach consensus on the importance of the preliminary list of different outcomes used in SIV VE studies. We thought *a priori* to include two rounds of voting followed by a final consensus meeting. Two Delphi rounds took place between July and September 2023. Although there is not a recognized sample size for conducting Delphi studies, a minimum of 12 panelists is generally considered acceptable to fulfill meaningful convergence [20]. By assuming a non-response rate of 25%, a total of 16 participants were invited to take part in Delphi through a formal personal written invitation. Experts were nominated to represent a diversified group of nationally recognized medical doctors with an expertise either in research on SIIV VE and/or in managing influenza vaccination in terms of providing advice to decision-makers or programming, organizing, and delivering vaccination. Experts were selected in the field of public health and preventive medicine ($n = 9$), family medicine ($n = 2$), infectious diseases ($n = 2$), internal medicine ($n = 1$), pediatrics ($n = 1$), and gerontology ($n = 1$). The imbalance across groups is justified by the fact that health professionals entrusted to manage the vaccination campaign in Italy are public health practitioners. Anyhow, physicians were also involved in considering the perspectives of health professionals more involved in vaccination offers in several settings.

Experts who agreed to participate were invited to the first kickoff virtual meeting, in which panelists were provided with problems, study goals, methods, timelines, and all major issues were interactively discussed. They were then sent an e-mail with a direct link to the questionnaire (File S1), which included relevant background information and instructions. The survey was performed on the virtual platform within3.com and was pre-tested for comprehension, overall flow, and any technical issues. Links to each round were active for 3 weeks and reminders were eventually sent.

Panelists were asked to score the outcomes in all given population target group on a 9-point anchored Likert scale, which consisted of a numeric range from 1 (least important) to 9 (most important). The 9-point anchored Likert scale was chosen based on the GRADE (grading of recommendations, assessment, development, and evaluations) guidelines [21]. Panelists were instructed to score each outcome individually without attempting to rank them. At the first round, panelists were also asked to comment on the questionnaire and to suggest additional outcomes or to modify the proposed ones.

Following the completion of the first round, experts' rankings were analyzed and expressed as medians with interquartile ranges (IQRs). Responses were a priori labeled according to the following categories: (i) 'critically important' for median scores 9 to 7; (ii) 'important, but not critical' for median scores 6 to 4 and (iii) 'not important' for median scores 3 to 1, as per available guidelines [21].

Target-specific outcomes that reached consensus (see below) by the end of round 1 were not brought to round 2. For round 2, panelists received both a summary of their previous responses and median (IQR) scores for each target-specific outcome that did not reach consensus. Panelists were then invited to re-rate the outcomes without knowing other panelists' scores as they only received the aggregated results of round 1.

Considering that no consensus measurement for Likert-based items is universally accepted [22], we used a priori a double criterion for defining consensus. For the first criterion, at least 75% of the expert rankings had to fall within one of the three above mentioned categories (i.e. critical, important, or not important). For the second criterion, the IQR had to be ≤ 2 [22–24]. Consensus was judged to be achieved if both criteria were satisfied.

All panelists were finally invited to join the final virtual face-to-face consensus meeting held in October 2023. This meeting aimed to discuss the findings and obtain the final rankings. In particular, participants were provided with rankings for those outcomes for which consensus was reached after both Delphi rounds. Outcomes that did not reach consensus were then shown and a group discussion on their importance was raised. Experts were then asked to attribute these outcomes to the final GRADE category, and the group discussion continued till unanimity was achieved. With respect to this last step, a preliminary proposal of classification was made, and inputs of the experts were collected. No substantial divergences emerged during the discussion.

3. Results

3.1. Literature review

A total of 1,210 records were initially identified, and the entire selection process is depicted in Figure S1. Briefly, following the removal of duplicates ($n = 204$), titles, and abstracts of 1,006 records were screened for their eligibility. Of these, 879 records were excluded based on title/abstract, while 17 were

excluded with reasons (Table S2). Therefore, 110 SRs were included in the analysis (Table S3).

For most (55%, 60/110) SRs, both randomized controlled trials (RCTs) and observational studies were included, while 25% (28/110) and 20% (22/110) assessed only SIV efficacy (including only RCTs) and effectiveness (including only observational studies), respectively. As expected, SRs of RCTs focused mainly on laboratory-confirmed outcomes. Conversely, the list of outcomes in SRs of both RCTs and observational studies was generally longer. At the level of single studies, we noted that large register-based and industry-sponsored studies tended to use less specific non-laboratory-confirmed outcomes. With regard to SIV target groups, 23% (25/110), 10% (11/110), 4% (4/110), 13% (14/110), 31% (34/110), and 5% (6/110) of SRs targeted the general population, children, working-age adults (mainly healthcare workers), older adults, subjects with different co-morbidities and pregnant women and/or newborns, respectively. The remaining 15% (16/110) of SRs considered mixed population groups. Compared with other target groups, SRs on pediatric populations were more frequently focused on LCI. There was a clear increasing trend in the number of published SRs: 75% (82/110) were published from 2015 onwards, while only one (1%) was published before 2000.

3.2. Selection and mapping of outcomes

A total of 489 pre-specified outcomes were extracted from the SRs included in the umbrella review (Table S3). Influenza-related outcomes with no or unknown laboratory confirmation and laboratory-confirmed outcomes were the most frequent and accounted for 27% (134/489) and 20% (100/489) of all outcomes, respectively. A total of 14% (69/489) and 13% (64/489) of outcomes were classified as hospital encounters and mortality outcomes not directly related to influenza, respectively. Outcomes relative to respiratory and cardiovascular complications potentially triggered by influenza virus counted up to 11% (55/489) and 8% (37/489), respectively. The remaining 6% (30/489) of outcomes were less frequent and formed a miscellaneous group (Table S3).

Following the data reduction process, 20 different outcomes were identified. As shown in Table 1, the number of outcomes ranged from eight in pregnant women to 18 in individuals with co-morbidities for a total of 63 outcomes across all target groups to be assessed by the panelists. LCI, ILI, influenza-related hospital encounters, and all-cause mortality were present in all target groups. Conversely, some outcomes were specific to a particular target group [e.g. asthma-related outcomes/recurrent wheezing for children and visits for respiratory causes, major cardiovascular events, all-cause intensive care unit (ICU) admission, and hospital encounters for exacerbation of preexisting health conditions for subjects with co-morbidities].

3.3. Delphi study

Of 16 experts invited, 14 agreed to participate (response rate 88%) and completed round 1. Their primary medical specialties were distributed as follows: public health and preventive medicine ($n = 8$), family medicine ($n = 2$), infectious diseases ($n = 1$), internal medicine ($n = 1$), pediatrics ($n = 1$), and gerontology ($n = 1$). They equally represented the academia ($n = 7$) and the practitioners/physicians ($n = 7$), and they were based in diverse macro-regions (Northern Italy, $n = 3$; Central Italy, $n = 8$; Southern Italy and Islands; $n = 3$). In this respect, it is important to report that influenza vaccination policies across Italian macro-regions do not change. A total of 882 single ranks were analyzed. No additional outcomes were suggested at round 1. Results of the first round were subject to the ceiling effect, i.e. 26% (230/882) of scores received the highest possible score equal to '9 – the most important.' As shown in Table 2, at round 1, consensus was reached for 35 out of 63

Table 1. Influenza-related outcomes used for the Delphi process by the target population group.

Outcome group	Outcome	Children	Working-age adults	Older adults	Subjects with co-morbidities	Pregnant women[a]
1	Any LCI	+	+	+	+	+
1	Hospital encounters for LCI	+	+	+	+	–
2	ARI	+	+	+	+	–
2	ILI	+	+	+	+	+
2	Influenza-related office visit	+	+	+	+	-
2	Influenza-related hospital encounter	+	+	+	+	+
2	Influenza-related mortality	+	+	+	+	–
2	Office visits for respiratory causes	–	–	–	+	–
3	Asthma-related outcomes/recurrent wheezing	+	–	–	–	–
3	Otitis media	+	+	–	–	–
3	Pneumonia and LRTD	+	+	+	+	–
3/4	Exacerbation of health conditions	–	+	–	+	–
4	Major cardiovascular events	–	–	–	+	–
5	All-cause hospital encounter	+	+	+	+	–
5	All-cause ICU admission	–	–	–	+	–
5	Hospital encounter for respiratory causes	–	–	+	+	–
5	Hospital encounter for cardiovascular causes	–	–	+	+	–
5	Hospital encounter for exacerbation of health conditions	–	–	–	+	–
6	All-cause mortality	+	+	+	+	+
6	Mortality for cardio-respiratory causes	–	–	+	+	–

[a]In this target population, all outcomes were distinguished between pregnant women and newborns.
ARI, acute respiratory infection; ICU, intensive care unit; ILI, influenza-like illness; LCI, laboratory confirmed influenza; LRTD, lower respiratory tract disease.

Table 2. Results of the Delphi process, by round and target population group.

	Round 1		Round 2	
Outcome	% consensus	Median (IQR)	% consensus	Median (IQR)
Children				
Influenza-related mortality	*93*	*9 (8.25–9)*	–	–
Hospital encounters for LCI	*86*	*9 (8–9)*	–	–
Pneumonia and LRTD	*79*	*8 (7.25–9)*	–	–
ARI	*79*	*7.5 (7–8.75)*	–	–
Influenza-related hospital encounter	71	7.5 (6.25–8)	*86*	*8 (7.25–8)*
All-cause mortality	57	7 (5.25–8)	71	7 (6.25–7)
All-cause hospital encounter	57	7 (6–7.75)	50	6.5 (6–7)
Any LCI	57	7 (4.5–9)	*86*	*7 (7–8)*
Otitis media	57	7 (5–7)	71	7 (6.25–7)
Influenza-related office visit	43	6.5 (5–8)	*79*	*7 (7–7)*
Asthma-related outcomes/recurrent wheezing	43	6.5 (6–7)	71	7 (6.25–7)
ILI	36	6.5 (4–7)	64	7 (6–7)
Working-age adults				
Hospital encounters for LCI	*100*	*9 (8–9)*	–	–
Influenza-related mortality	*100*	*9 (8–9)*	–	–
Exacerbation of health conditions	*93*	*8 (7.25–9)*	–	–
Influenza-related hospital encounter	*93*	*7.5 (7–8)*	–	–
Pneumonia and LRTD	*86*	*8 (7–8.75)*	–	–
Any LCI	71	8.5 (5.5–9)	*93*	*9 (8.25–9)*
All-cause hospital encounter	71	7 (6.25–7.75)	*86*	*7 (7–7)*
All-cause mortality	71	7 (6.25–8)	*79*	*7 (7–7)*
ARI	69	7 (6–8)	*93*	*7 (7–8)*
Influenza-related office visit	57	7 (5.25–7.75)	*93*	*7.5 (7–8)*
Otitis media	50	4.5 (3–6)	*86*	*4 (4–5)*
ILI	36	6.5 (6–7)	64	7 (6–7)
Older adults				
Influenza-related mortality	*100*	*9 (8.25–9)*	–	–
Hospital encounters for LCI	*100*	*8.5 (8–9)*	–	–
Mortality for cardio-respiratory causes	*93*	*8 (8–8.75)*	–	–
Pneumonia and LRTD	*93*	*8 (7–9)*	–	–
Influenza-related hospital encounter	*93*	*7 (7–8)*	–	–
Hospital encounter for respiratory causes	*86*	*8 (7.25–8)*	–	–
Hospital encounter for cardiovascular causes	*86*	*8 (7–8)*	–	–
All-cause hospital encounter	*79*	*7 (7–8)*	–	–
All-cause mortality	*79*	*7 (7–8)*	–	–
ARI	74	7 (6.25–8)	*93*	*7 (7–8)*
Any LCI	71	8.5 (6.25–9)	*96*	*9 (8–9)*
Influenza-related office visit	64	7 (6–8)	*93*	*8 (7–8)*
ILI	43	6.5 (5.25–7.75)	*86*	*7 (7–7)*
Subjects with co-morbidities				
Influenza-related mortality	*100*	*9 (8–9)*	–	–
Pneumonia and LRTD	*100*	*8 (8–9)*	–	–
Hospital encounter for exacerbation of health conditions	*100*	*8 (8–8.75)*	–	–
Hospital encounter for respiratory causes	*100*	*8 (7.25–8.75)*	–	–
Mortality for cardio-respiratory causes	*93*	*8 (7.25–8.75)*	–	–
Influenza-related hospital encounter	*93*	*8 (7–8)*	–	–
Major cardiovascular events	*93*	*8 (7–8)*	–	–
All-cause ICU admission	*93*	*7 (7–8)*	–	–
Hospital encounters for LCI	*86*	*8.5 (8–9)*	–	–
Exacerbation of health conditions	*86*	*8 (7–8)*	–	–
Hospital encounter for cardiovascular causes	*79*	*7.5 (7–8)*	–	–
ARI	*79*	*7 (7–8)*	–	–
Any LCI	71	8 (6.25–9)	*100*	*9 (8–9)*
All-cause hospital encounter	71	7 (6.25–7)	*100*	*7 (7–7)*
All-cause mortality	64	7 (6–7.75)	*86*	*7 (7–7)*
Influenza-related office visit	57	7 (5–8)	*93*	*7.5 (7–8)*
ILI	57	7 (5–7)	*86*	*7 (7–8)*
Office visits for respiratory causes	35	6.5 (5–7)	71	7 (6.25–7)
Pregnant women				
All-cause maternal mortality	*100*	*9 (7.25–9)*	–	–
Influenza-related hospital encounter in pregnant woman	*93*	*8 (7–9)*	–	–
Influenza-related hospital encounter in newborn	*86*	*9 (8–9)*	–	–
All-cause neonatal mortality	*86*	*9 (7–9)*	–	–
Any LCI in pregnant woman	*79*	*7.5 (7–9)*	–	–
Any LCI in newborn	71	9 (6.25–9)	*100*	*9 (9–9)*
ILI in newborn	64	7 (6–8)	*93*	*7 (7–7)*
ILI in pregnant woman	43	6.5 (6–8)	*86*	*7 (7–7)*

ARI, acute respiratory infection; ICU, intensive care unit; ILI, influenza-like illness; LCI, laboratory confirmed influenza; LRTD, lower respiratory tract disease. Outcomes for which consensus was reached are evidenced in italics.

outcomes (56%). All these outcomes were judged critically important. Regarding specific SIV target groups, most outcomes for older adults (69%; 9/13), subjects with co-morbidities (67%; 12/18) and pregnant women (63%; 5/8) reached consensus at round 1. Conversely, only 33% (4/12) and 42% (5/12) of the outcomes specific to children and working-age adults, respectively, reached agreement at round 1. On considering median votes, in all target groups, experts tended to attribute the highest ranks to more severe outcomes, such as influenza-related mortality and hospital encounters for LCI. By contrast, syndrome-based and primary care outcomes like ILI or influenza-related office visits with no laboratory confirmation were generally scored as less important. Analogously, as shown by the width of IQRs, ratings for the target groups of older adults and subjects with co-morbidities were the most consistent. Less consistency was instead observed for the outcomes relative to LCI and syndromic definitions, especially in children (Table 2).

At round 2, no attrition was detected as all the panelist involved in round 1 also completed round 2, and consensus was reached for all the remaining outcomes in older adults and pregnant women, while in the target groups of working-age adults and subjects with co-morbidities consensus was achieved for all but one outcome (ILI and office visits for respiratory causes, respectively). In children, consensus was not reached for the outcomes relative to all-cause mortality and hospital encounters, otitis media, ILI, and asthma.

Finally, following the presentation of results of the first two Delphi rounds and group discussion during the consensus meeting, all participants agreed that the outcomes of ILI (children and working-age adults), office visits for respiratory causes in subjects with co-morbidities, all-cause mortality, all-cause hospital encounters, otitis media, and asthma-related outcomes in children should be considered as important but not critical.

4. Discussion

In recent years, observational studies on absolute and, especially, relative SIV VE investigated a plethora of heterogeneous outcomes. This heterogeneity underlines the need for a core set of outcome measures for the synthesis of relevant evidence in light of developing vaccination recommendations. To our knowledge, this is the first study that attempted to map, organize, and rank this multitude of outcomes used in SIV VE studies.

Our umbrella review highlighted a large variability of study outcomes, especially in some target groups such as subjects with underlying health conditions. This variability is primarily driven by observational studies that may use both high (e.g. SARI with LCI) and low (e.g. all-cause mortality) specificity outcomes. Conversely, RCTs are usually based on highly specific outcomes, namely LCI. It is axiomatic that as the specificity of an outcome increases, its sensitivity decreases and *vice versa* [25]. Less specific outcomes may address the issue of underutilization of laboratory diagnosis of influenza and thus allow evaluating the hidden burden of influenza [8].

The variability of the considered outcomes also increased over the past few years. One of the possible reasons for this expansion is an increased availability of and interest in real-world data derived from large electronic health and medical claims records. It appears that most available real-world evidence studies have been conducted or committed by industry, and their SIV VE estimates are mostly based on less specific non-laboratory-confirmed outcomes [18,26]. Indeed, both public [27] and private [28] registries usually lack access to virological case confirmation data. Considering a continuously evolving SIV market, it is likely that VE studies based on real-world data with no LCI data will increase. Our results may therefore guide decisions on the selection of the most relevant primary and secondary outcomes in future SIV VE studies and help assessors to weigh VE estimates with respect to the different outcomes.

We then showed that the variety of outcomes used is also explained by population target groups. There was a general tendency to consider more severe outcomes with increasing age and the presence of co-morbidities. Furthermore, in the latter target group, there was also a wide interest in investigating the effect of influenza infection on outcomes related to the underlying health condition.

According to the World Health Organization (WHO) [3], for the purpose of assessing SIV VE, outcomes may be dichotomized according to two major attributes, namely specificity for influenza virus infection (being LCI the most specific outcome) and disease severity (ranging from ambulatory visits for mild disease to fatal cases). Our Delphi study showed that more severe outcomes had on average higher scores and often independently from laboratory confirmation. For instance, influenza-related mortality with no explicit virological diagnosis was the top-ranked endpoint in children, older adults, and subjects with co-morbidities. On the other hand, the outcome 'any LCI,' which is usually used in RCTs for measuring SIV efficacy, ranked high in all target groups except children. Braunfeld et al. [29] have reported that the majority of pediatric SIV RCTs focused on capturing any influenza illness, while only a few included other clinically relevant outcomes, such as hospitalization for LCI, all-cause hospitalization, LCI with ICU admission, or serious extra-pulmonary complications [29]. It is noteworthy that disease severity may have a significant effect on SIV efficacy even in RCTs. For example, in a pediatric RCT [30] the per-protocol efficacy of a quadrivalent SIV was 55.4% and 73.1% against any LCI and moderate-to-severe (fever ≥39°C, otitis media, LRTD, or serious extrapulmonary complications) LCI, respectively, with a 32% relative difference. Analogously, a six-season observational study by Godoy et al. [31] has demonstrated that SIV reduced the severity of disease even in cases where it did not prevent infection *per se*. In this regard, the WHO has called [32] for a larger adoption of standardized, and ideally validated, severe influenza-related illness outcomes (e.g. SARI). Indeed, SIV VE against severe illness is expected to be of higher public health value and to have a greater impact on policy-making and the findings of our study are aligned with that. Conversely, VE against non-severe LCI can be more relevant for target groups and settings, in which the herd protection is important (e.g. healthcare workers) [32].

Outpatient syndromic-based outcomes like ILI and ARI received comparatively low rankings in all target groups.

Apart from being non-severe endpoints, other reasons may have contributed to this finding. First, while these syndromes are frequently used as triggers for specimen collection in both RCTs [29] and observational studies [33], their operational definitions and associated predictive values vary [34]. Second, ILI/ARI may be caused by tens of different viruses and bacteria and the underlying etiology changes from season to season [35]. The recent availability of COVID-19 and respiratory syncytial virus vaccines [36] (both infections share several signs and symptoms with influenza) may have further downgraded the value of syndrome-based outcomes.

In our Delphi study, several outcomes were judged to be critically important, and their median ranks were very close. Inclusion of different outcomes in a VE study may therefore provide more insights into a spectrum of SIV benefits and possibly satisfy different expectations of different stakeholders. Indeed, the WHO guidelines have also underlined [3] that the selection of a single outcome (e.g. pneumonia) may underestimate the overall SIV benefits since other important outcomes that influenza vaccine may prevent (e.g. exacerbation of preexisting diseases and cardiovascular disease) are overlooked. Furthermore, the importance of outcomes could also depend on the considered stakeholder: researchers and medical doctors may have different preferences as compared to decision-makers and regulators, while patients may still prefer other outcomes. Also, even within the same group of stakeholders, including medical doctors, different preferences may emerge. Indeed, our work was aimed at capturing medical doctors' perspectives. This is a crucial starting point considering that any health intervention needs to be first effective and safe before being considered for use in the whole population and that medical doctors are those entrusted to vaccinate people. Nevertheless, it is important to point out that it is also crucial to collate different stakeholders' perspectives to more comprehensively support the decision-making [37].

This study may suffer from some limitations. First, we must acknowledge that some potentially relevant outcomes may have been omitted. On the other hand, no additional outcomes have been suggested by the panelists. Nevertheless, we believe that any potential missed relevant outcome would be only a subcategory of already included ones (e.g. outcomes with a slightly different operational definition) and therefore would have the same or very close rank. Indeed, the inclusion of all available outcome definitions was unrealistic for the Delphi process. Second, we noted some ceiling effect, in which panelists tended to attribute higher ranks to all outcomes. For the same reason, none of the proposed outcomes was deemed 'not important.' A similar effect has been reported in a study, in which Delphi panelists were randomized to rate outcomes on either a 3-point or a 9-point Likert scale [23]. Compared with the 3-point scale, the use of the 9-point scale resulted in twice as many outcomes being rated as 'critically important.' Third, our Delphi process involved only Italian experts, and therefore its results may not be fully transferable to other realities. For instance, in regions with a high usage of laboratory confirmation (e.g. a large use of point-of-care testing in primary care), LCI-related outcomes could have a higher importance. Finally, our panelists were chosen among experts with a medical background, therefore our work does not grasp all the stakeholders' perspectives. In order to be considered as 'core outcome set' (COS) according to the Core Outcome Measures in Effectiveness Trials Initiative (COMET) [38] outcomes should also be scored by other groups of stakeholders, namely patients and carers. Nevertheless, it should be considered that the inclusion of patients and carers would be rather challenging in terms of the credibility and utility of COS, because of the specific characteristics of the disease and the intervention, i.e. vaccination.

5. Conclusion

In conclusion, over the last decade studies investigating the absolute and relative SIV VE used numerous outcomes that differ significantly in terms of their predictive ability. Owing to both an increasing interest in real-world evidence and large underutilization of laboratory confirmation, the use of less specific outcomes in VE research will likely increase. The abundance of different outcomes, however, may generate contrasting results and therefore confound decision-making. This Delphi study ranked the outcomes used in SIV VE studies and showed that the Italian medical community attaches greater importance to more severe influenza-related outcomes, independently of the fact whether they are laboratory-confirmed or not. Our results may constitute the ground for the future establishment of a core outcome set to be recommended for inclusion in protocols of SIV VE studies, which would allow for more direct between-study comparisons. Finally, our findings could be useful to support the decision-making process leading to influenza vaccination recommendations.

Funding

This work was funded by Seqirus S.r.L., a pharmaceutical company that manufactures and commercializes influenza vaccines. The study funder had no role in study design and conduct but, as per contract, could review the manuscript draft by making non-mandatory suggestions.

Declaration of interest

A Domnich was previously a permanent employee of Seqirus S.r.L. Both A Domnich and C de Waure received a fee from Seqirus S.r.L. for coordinating the project. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or material discussed in the manuscript apart from those disclosed.

Author contributions

A Domnich and C de Waure were involved in the conception, design, analysis, and interpretation of data; all the other authors were involved in the acquisition and interpretation of data. AD and CdW drafted the paper, and all the other Authors revised it for intellectual content and provided the final approval. All the authors agree to be accountable for all aspects of the work.

Reviewer disclosures

A reviewer on this manuscript has received an honorarium for their review work. Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

ORCID

Chiara de Waure http://orcid.org/0000-0002-4346-1494
Vincenzo Baldo http://orcid.org/0000-0001-6012-9453
Paolo Bonanni http://orcid.org/0000-0003-2875-3744
Michele Conversano http://orcid.org/0000-0002-5245-0340
Alberto Fedele http://orcid.org/0000-0003-1198-7042
Giovanni Gabutti http://orcid.org/0000-0002-7259-4923
Roberto Ieraci http://orcid.org/0009-0001-1494-055X
Francesco Landi http://orcid.org/0000-0002-3472-1389
Andrea Orsi http://orcid.org/0000-0002-2433-9610
Caterina Rizzo http://orcid.org/0000-0002-5583-7508
Alberto Villani http://orcid.org/0000-0002-9120-0424
Francesco Vitale http://orcid.org/0000-0002-6354-8060
Alexander Domnich http://orcid.org/0000-0001-9608-8892

References

1. World Health Organization (WHO). Vaccines against influenza: WHO position paper – May 2022. Wkly Epidemiol Rec. 2022;97 (19):185–208.
2. Kim DeLuca E, Gebremariam A, Rose A, et al. Cost-effectiveness of routine annual influenza vaccination by age and risk status. Vaccine. 2023;41(29):4239–4248. doi: 10.1016/j.vaccine.2023.04.069
3. World Health Organization (WHO). Evaluation of influenza vaccine effectiveness: a guide to the design and interpretation of observational studies. Geneva: WHO; 2017 [cited 2023 Oct 16]. Available from: https://apps.who.int/iris/bitstream/handle/10665/255203/9789241512121-eng.pdf
4. Belongia EA, Simpson MD, King JP, et al. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. Lancet Infect Dis. 2016;16(8):942–951. doi: 10.1016/S1473-3099(16)00129-8
5. Domnich A, Manini I, Calabrò GE, et al. Mapping host-related correlates of influenza vaccine-induced immune response: an umbrella review of the available systematic reviews and meta-analyses. Vaccines (Basel). 2019;7(4):215. doi: 10.3390/vaccines7040215
6. Simonsen L, Taylor RJ, Viboud C, et al. Mortality benefits of influenza vaccination in elderly people: an ongoing controversy. Lancet Infect Dis. 2007;7(10):658–666. doi: 10.1016/S1473-3099(07)70236-0
7. Lapi F, Domnich A, Marconi E, et al. Adjuvanted versus non-adjuvanted standard-dose influenza vaccines in preventing all-cause hospitalizations in the elderly: a cohort study with nested case-control analyses over 18 influenza seasons. Expert Rev Vaccines. 2022;21(11):1647–1653. doi: 10.1080/14760584.2022.2115362
8. Sellers SA, Hagan RS, Hayden FG, et al. The hidden burden of influenza: a review of the extra-pulmonary complications of influenza infection. Influenza Other Respir Viruses. 2017;11(5):372–393. doi: 10.1111/irv.12470
9. Centers for Disease Control and Prevention (CDC). Guide for considering influenza testing when influenza viruses are circulating in the community. 2020 [cited 2023 Oct 16]. Available from: https://www.cdc.gov/flu/professionals/diagnosis/consider-influenza-testing.htm
10. López-Medrano F, Alfayate S, Carratalà J, et al. Executive summary. Diagnosis, treatment and prophylaxis of influenza virus infection. Consensus statement of the Spanish society of infectious diseases and clinical microbiology (SEIMC), the Spanish society of pediatric infectious diseases (SEIP), the Spanish association of vaccinology (AEV), the Spanish society of family and community medicine (SEMFYC) and the Spanish society of preventive medicine, public health and health management (SEMPSPGS). Enferm Infecc Microbiol Clin (Engl Ed). 2023;41(2):111–122. doi: 10.1016/j.eimce.2022.10.013
11. Orenstein EW, De Serres G, Haber MJ, et al. Methodologic issues regarding the use of three observational study designs to assess influenza vaccine effectiveness. Int J Epidemiol. 2007;36 (3):623–631. doi: 10.1093/ije/dym021
12. Jackson ML, Rothman KJ. Effects of imperfect test sensitivity and specificity on observational studies of influenza vaccine effectiveness. Vaccine. 2015;33(11):1313–1316. doi: 10.1016/j.vaccine.2015.01.069
13. Ferdinands JM, Gargiullo P, Haber M, et al. Inactivated influenza vaccines for prevention of community-acquired pneumonia: the limits of using nonspecific outcomes in vaccine effectiveness studies. Epidemiol. 2013;24(4):530–537. doi: 10.1097/EDE.0b013e3182953065
14. Morris SE, Grohskopf LA, Ferdinands JM, et al. Evaluating potential impacts of a preferential vaccine recommendation for adults 65 years of age and older on US influenza burden. Epidemiol. 2023;34 (3):345–352. doi: 10.1097/EDE.0000000000001603
15. Lewis NM, Chung JR, Uyeki TM, et al. Interpretation of relative efficacy and effectiveness for influenza vaccines. Clin Infect Dis. 2022;75(1):170–175. doi: 10.1093/cid/ciab1016
16. Coleman BL, Sanderson R, Haag MDM, et al. Effectiveness of the MF59-adjuvanted trivalent or quadrivalent seasonal influenza vaccine among adults 65 years of age or older, a systematic review and meta-analysis. Influenza Other Respir Viruses. 2021;15 (6):813–823. doi: 10.1111/irv.12871
17. Lee JKH, Lam GKL, Shin T, et al. Efficacy and effectiveness of high-dose influenza vaccine in older adults by circulating strain and antigenic match: an updated systematic review and meta-analysis. Vaccine. 2021;39(Suppl 1):A24–A35. doi: 10.1016/j.vaccine.2020.09.004
18. Domnich A, de Waure C. Comparative effectiveness of adjuvanted versus high-dose seasonal influenza vaccines for older adults: a systematic review and meta-analysis. Int J Infect Dis. 2022;122:855–863. doi: 10.1016/j.ijid.2022.07.048
19. Belbasis L, Bellou V, Ioannidis JPA. Conducting umbrella reviews. BMJ Med. 2022;1(1):e000071. doi: 10.1136/bmjmed-2021-000071
20. Murphy MK, Black NA, Lamping DL, et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess. 1998;2(3):i–iv,1–88. doi: 10.3310/hta2030
21. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol. 2011;64(4):395–400. doi: 10.1016/j.jclinepi.2010.09.012
22. von der Gracht Ha, von der Gracht HA. Consensus measurement in Delphi studies: review and implications for future quality assurance. Technol Forecasting Social Change. 2012;79 (8):1525–1536. doi: 10.1016/j.techfore.2012.04.013
23. De Meyer D, Kottner J, Beele H, et al. Delphi procedure in core outcome set development: rating scale and consensus criteria determined outcome selection. J Clin Epidemiol. 2019;111:23–31. doi: 10.1016/j.jclinepi.2019.03.011
24. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of delphi studies. J Clin Epidemiol. 2014;67(4):401–409. doi: 10.1016/j.jclinepi.2013.12.002
25. Carter JV, Pan J, Rai SN, et al. ROC-ing along: evaluation and interpretation of receiver operating characteristic curves. Surgery. 2016;159(6):1638–1645. doi: 10.1016/j.surg.2015.12.029
26. European Medicines Agency (EMA). Real-world evidence framework to support EU regulatory decision-making. 2023 [cited 2023 Oct 16]. Available from: https://www.ema.europa.eu/en/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-report-experience-gained_en.pdf
27. Izurieta HS, Lu M, Kelman J, et al. Comparative effectiveness of influenza vaccines among US medicare beneficiaries ages 65 years and older during the 2019–2020 season. Clin Infect Dis. 2021;73 (11):e4251–e4259. doi: 10.1093/cid/ciaa1727
28. Pelton SI, Divino V, Postma MJ, et al. A retrospective cohort study assessing relative effectiveness of adjuvanted versus high-dose trivalent influenza vaccines among older adults in the United States during the 2018–19 influenza season. Vaccine. 2021;39 (17):2396–2407. doi: 10.1016/j.vaccine.2021.03.054
29. Braunfeld JB, Carson HN, Williams SR, et al. Clinical endpoints to inform vaccine policy: a systematic review of outcome measures from pediatric influenza vaccine efficacy trials. Vaccine. 2022;40 (32):4339–4347. doi: 10.1016/j.vaccine.2022.06.028
30. Jain VK, Rivera L, Zaman K, et al. Vaccine for prevention of mild and moderate-to-severe influenza in children. N Engl J Med. 2013;369 (26):2481–2491. doi: 10.1056/NEJMoa1215817

31. Godoy P, Romero A, Soldevila N, et al. Influenza vaccine effectiveness in reducing severe outcomes over six influenza seasons, a case-case analysis, Spain, 2010/11 to 2015/16. Euro Surveill. 2018;23 (43):1700732. doi: 10.2807/1560-7917.ES.2018.23.43.1700732
32. World Health Organization (WHO). WHO preferred product characteristics for next-generation influenza vaccines. 2017 [cited 2023 Oct 16]. https://apps.who.int/iris/bitstream/handle/10665/258767/9789241512466-eng.pdf
33. Fukushima W, Hirota Y. Basic principles of test-negative design in evaluating influenza vaccine effectiveness. Vaccine. 2017;35 (36):4796–4800. doi: 10.1016/j.vaccine.2017.07.003
34. Casalegno JS, Eibach D, Valette M, et al. Performance of influenza case definitions for influenza community surveillance: based on the French influenza surveillance network GROG, 2009-2014. Euro Surveill. 2017;22(14):30504. doi: 10.2807/1560-7917.ES.2017.22.14.30504
35. Thomas RE. Is influenza-like illness a useful concept and an appropriate test of influenza vaccine effectiveness? Vaccine. 2014;32(19):2143–2149. doi: 10.1016/j.vaccine.2014.02.059
36. Grohskopf LA, Blanton LH, Ferdinands JM, et al. Prevention and control of seasonal influenza with vaccines: recommendations of the advisory committee on immunization practices — United States, 2023–24 influenza season. MMWR Recomm Rep. 2023;72 (RR–2):1–25. doi: 10.15585/mmwr.rr7202a1
37. Thorlund K, Smith D, Linsell C, et al. The importance of appropriate selection of clinical endpoints in outpatient COVID-19 clinical trials. Commun Med (Lond). 2023;3(1):53. doi: 10.1038/s43856-023-00281-1
38. Williamson PR, Altman DG, Bagley H, et al. The COMET handbook: version 1.0. Trials. 2017;18(Suppl 3):280. doi: 10.1186/s13063-017-1978-4