Cochrane re-arranged: Support for policies to vaccinate elderly people against influenza

Walter E.P. Beyer a, Janet McElhaney b, Derek J. Smith c,d,a, Arnold S. Monto e, Jonathan S. Nguyen-Van-Tam f, Albert D.M.E. Osterhaus a,∗

a Erasmus MC, Department of Viroscience, PB 2040, 3000 CA Rotterdam, The Netherlands
b University of Connecticut Health Center, Department of Immunology, 263 Farmington Avenue, Farmington, CT 06030–1915, USA
c University of Cambridge, Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK
d World Health Organization Collaborating Center for Modeling, Evolution, and Control of Emerging Infectious Diseases, Cambridge CB2 3EJ, UK
e University of Michigan, School of Public Health, 1415 Washington Heights, Ann Arbor, MI 48109–2025, USA
f University of Nottingham, Health Protection and Influenza Research Group, Division of Epidemiology and Public Health, Clinical Sciences Building, Hucknall Road, Nottingham NG5 1PB, UK

ARTICLE INFO

Article history:
Received 16 July 2013
Received in revised form 20 September 2013
Accepted 26 September 2013
Available online xxx

Keywords:
Efficacy
Effectiveness
Safety
Influenza
Vaccination
Elderly

ABSTRACT

The 2010 Cochrane review on efficacy, effectiveness and safety of influenza vaccination in the elderly by Jefferson et al. covering dozens of clinical studies over a period of four decades, confirmed vaccine safety, but found no convincing evidence for vaccine effectiveness (VE) against disease thus challenging the ongoing efforts to vaccinate the elderly.

However, the Cochrane review analyzed and presented the data in a way that may itself have hampered the desired separation of real vaccine benefits from inevitable ‘background noise’. The data are arranged in more than one hundred stand-alone meta-analyses, according to various vaccine types, study designs, populations, and outcome case definitions, and then further subdivided according to virus circulation and antigenic match. In this way, general vaccine effects could not be separated from an abundance of environmental and operational, non-vaccine-related variation. Furthermore, expected impacts of changing virus circulation and antigenic drift on VE could not be demonstrated.

We re-arranged the very same data according to a biological and conceptual framework based on the basic sequence of events throughout the ‘patient journey’ (exposure, infection, clinical outcome, observation) and using broad outcome definitions and simple frequency distributions of VE values. This approach produced meaningful predictions for VE against influenza-related fatal and non-fatal complications (average ~30% with large dispersion), typical influenza-like illness (~40%), disease with confirmed virus infection (~50%), and biological vaccine efficacy against infection (~60%), under conditions of virus circulation. We could also demonstrate a VE average around zero in the absence of virus circulation, and decreasing VE values with decreasing virus circulation and increasing antigenic drift.

We regard these findings as substantial evidence for the ability of influenza vaccine to reduce the risk of influenza infection and influenza-related disease and death in the elderly.

© 2013 The Authors. Published by Elsevier Ltd. All rights reserved.

1. The Cochrane review is challenging policies to vaccinate the elderly against influenza

If infected with seasonal influenza (virus type A or B), elderly people encounter a substantially elevated risk of developing non-fatal and fatal complications than adult persons of younger age [1]. Does influenza vaccination protect the elderly from influenza-related disease and death, and to what degree? A heated debate has recently arisen about this question. In particular, the comprehensive Cochrane review of 75 selected publications on influenza vaccine effectiveness (VE) and safety in the elderly [2] has been cited as a reason not to vaccinate the elderly, or at the very least to be highly sceptical about the possible public health benefits. Policies to vaccinate the elderly, as established in many countries [3], have been challenged and sometimes undermined.

Indeed, the Cochrane review concludes that the data from the selected original publications “are so biased as to be virtually uninterpretable”. Principal concerns relate to the scarcity of randomized
controlled trials (RCTs), the large number of observational studies of "very low quality" (with concomitant risk of selection bias), and inconsistent findings. The authors conclude that policies for vaccinating the elderly are either non-evidence-based or, at best, based on poor quality evidence, and therefore vaccination of the elderly is not encouraged; a placebo-controlled trial run over several seasons should instead decide the question.

We hesitate to accept that the research efforts already made over several decades, with hundreds of studies performed and millions of persons participating, "consistently fails to give satisfactory answers" and has led to nothing more than inconclusive results. We note that the Cochrane authors had arranged the included data in more than one hundred stand-alone strata, according to various vaccine types, study designs, populations, outcome case definitions, virus circulation and degree of antigenic matching. When analysing the data in this way, general vaccine effects might not be visible given an abundance of non-vaccine-related variation. In this case, it is potentially not the original studies that fail to give satisfactory answers, but the method in which they have been meta-analyzed.

Encouraged by the invitation of the Cochrane authors to produce "any alternative interpretation" of the evidence, we wondered how the very same data would behave when stratified according to a small number of soundly-based and disease-informed scenarios rather than in a blind meta-analysis textbook approach that has more than 100 scenarios. To be able to directly test the choice of stratification, we accepted verbatim the choices for inclusion and exclusion of publications made by the Cochrane authors, and used the exact same data. We approached our analysis from the basic sequence of events in the patient journey (exposure, infection, outcome, disease, and observation) to construct a simple biological and conceptual framework (Supporting Material 1). Accepted knowledge of the infection process [4], the rationale of vaccination [5], and a consideration of vaccine quality [6,7] were all taken into account. With only few, but broad, outcome definitions, we could express the results by simple frequency distributions of VE values, and corroborate them by the inverse variance-weighted method [8], which produces meta-analyzed VE means and associated confidence intervals.

Here, we show that our approach leads to meaningful predictions for VE against influenza-related fatal and non-fatal complications (average ~30% with large dispersion), typical influenza-like illness (ILI) without virus confirmation (~40%), disease with laboratory confirmation of virus infection (~50%), and biological vaccine efficacy against infection (at least 50%, more likely ~60%), under conditions of known virus circulation. Minor and major antigenic drift has only a small average impact. We regard these findings as substantial evidence for the ability of influenza vaccine to reduce the risk of influenza infection and consequently influenza-related disease and death in the elderly. In our view, existing policies to vaccinate the elderly are vindicated and should be further implemented worldwide.

2. The original publications comprise 40 years of clinical research and are inevitably heterogeneous – this can be positively exploited

In their latest update published in 2010 [2], the Cochrane authors excluded 255 of 330 retrieved candidate publications (77%), and presented detailed information about the 75 included publications. Safety issues regarding influenza vaccines were not raised, and accordingly these are not examined in our paper. We focussed on vaccine effectiveness (VE), i.e., the vaccine’s ability to prevent an observed clinical outcome. Eventually, 248 VE values from one hundred trial-seasons were available from the Cochrane tables. Studies were performed in community-dwellers and nursing home residents, and took place across four continents from 1965 to 2005. This period covers the A-H3N2 pandemic (1968) and the re-emergence of the A-H1N1 subtype (1977). Few trials were RCTs, while the vast majority of trials were observational (case-control trials and cohort trials).

Inevitably, these data are heterogeneous by accidental, not vaccine-related variation, i.e., the varying influence of individual, social, viral, seasonal, spatial, temporal and operational factors. Meta-analyses with an abundance of outcome definitions and formal strata may not be the optimal method to analyze these kinds of data. To unravel the real biological vaccine efficacy, i.e., the vaccine’s ability to prevent infection, the number of outcome definitions was substantially reduced from 14 to only three: laboratory-confirmed disease of any kind, influenza-like illness (ILI), and non-fatal and fatal complications (i.e., serious morbidity and death from influenza-related causes). The outcome ‘all-cause death’ was handled separately because of extremely low specificity (see below).

Fig. 1 shows how the VE values from large numbers of heterogeneous trials would be distributed, with similar impacts of positive and negative accidental factors assumed (details are given in Supporting Material 1). When virus circulates during a trial and the vaccine is efficacious (Fig. 1A), the VE distributions rank hierarchically from lowest average outcome specificity (complications, blue) to that with highest specificity (laboratory-confirmed disease, green); all together they point to biological vaccine efficacy (black), the principal preventive effect of vaccination, which is not assessed directly. In the case of vaccine failure (vaccine efficacy varying around zero, Fig. 1B), the VE distributions would all peak at zero. In this visual way, we could discriminate between two basic scenarios despite, or thanks to, large heterogeneity between the original studies.

3. Within a biological framework, Cochrane’s data reveal the efficacy of vaccinating the elderly

All we had to do was to group Cochrane’s VE values according to virus circulation and alternative outcome, and to inspect, whether the resulting distributions would fit with the scenarios in Fig. 1. Our data processing is described in Supporting Material 2. We first excluded a number of VE values, mostly to avoid data redundancy. Our final database consisted of 165 VE values from 95 trials, involving 2,504,162 elderly persons, and described in 64 publications. We identified 40 VE values from trials likely performed in absence of virus circulation. Another 13 VE values against ‘all-cause death’ were treated separately. Finally, we could examine the vaccine performance under virus circulation, from the remaining 112 VE values (Fig. 2).

The distributions clearly do not resemble Fig. 1B, the scenario of vaccine failure and zero biological vaccine efficacy. Instead, the resemblance with the patterns of Fig. 1A is striking. The VE values against complications peak between 30 and 50%, with a meta-analyzed mean of 28% (95% CI 26–30%). VE values against ILI have a broad maximum between 30 and 70%; their meta-analyzed mean is plausibly larger (39%; 35–43%) than the effect against complications. VE values against laboratory-confirmed disease show a narrow distribution, which peaks between 50 and 70% (49%; 33–62%). Visual examination of the three distributions and comparison with those in Fig. 1 let us predict the biological vaccine efficacy in this dataset as being at least 50%, but likely 60% or larger.

In our approach, the data produce plausible VE patterns with respect to degree of virus circulation (decreasing VE values with decreasing virus circulation, see Supporting Material 2). Moreover, we found two interesting time trends, not noticed by the
Cochrane authors: The average vaccine coverage substantially rose from 43% in 1971 to 64% in 2004, and the average VE substantially fell from 52% to 26% in the same period. We discuss whether these time trends maybe causally related. Annually increasing vaccine coverage may successively hamper virus transmission and reduce exposure risk. This beneficial mass treatment effect would have a paradoxically negative impact on the VE measure, which decreases with decreasing exposure risk, falsely suggesting a decline in vaccine benefits.

The VE patterns with respect to antigenic drift are also plausible: VE values decrease with increasing antigenic distance between vaccine components and circulating strains. It is worth noting that even under conditions of substantial antigenic drift, vaccine still provided considerable protection, a finding not presented in the Cochrane review.

The outcome ‘all-cause death’ was analyzed separately. Given a low outcome specificity of up to 10%, VE values against this outcome would be expected to peak between zero and 5%. The distribution of the 13 all-cause death values, however, showed a surprising maximum between 50 and 70%, and a meta-analyzed mean of 48% (95% CI 47–50%) suggesting an overriding ‘healthy user’ bias. According to Simonsen et al. [9], the combination of both low outcome specificity and bias “has produced a high degree of mismeasurement, leading to greatly inflated estimates”. Simonsen advises against using this outcome for the assessment of VE. We agree, certainly for the studies involved here. Interestingly, Fireman et al. [10] applied a new ingenious manner of bias control to all-cause mortality data from a large elderly population during nine influenza seasons. They found a plausible VE estimate of 4.6% (95% CI 0.7–8.3%) during influenza season, while the average excess mortality was estimated at 7.8%.

4. Why did the Cochrane review not reach similar conclusions?

The Cochrane authors arranged the 248 identified VE values into 15 main comparisons with more than 100 different strata. Furthermore, there is an abundance of fourteen outcome definitions. We would call this arrangement an over-stratification, which has impeded the identification of vaccine effects. Their analysis was further weakened by disturbing misclassification. We could demonstrate, for example, that Cochrane’s sub-division for the dichotomy ‘outbreak/epidemic’ versus ‘no outbreak/no epidemic’ was not successful. Some trials performed under virus circulation as described in the original publications, were nevertheless classified as ‘no outbreak/no epidemic’, and vice versa (Supporting Material 3).

In conclusion, the Cochrane review usefully collects data from a large number of studies on the effects of influenza vaccination in the elderly, but Cochrane’s data analysis is not guided by biological criteria and has consequently regarded the data as being inconclusive and uninterpretable. In contrast, our re-arrangement of the same data into more clinically meaningful scenarios has demonstrated the beneficial effects of vaccination. We suggest that vaccination of the elderly is efficacious in reducing infection, disease, and death, caused by influenza virus infection; is worthwhile as a public health intervention; and that there is a sound scientific basis for the recommendations made by the World Health Organization, and multiple international and national bodies.

Certainly, the Cochrane authors have a point when criticizing short-cut messages to the public as “flu shot prevents death”. Such misleading messages are based on biased ‘all-cause death’ studies and should be avoided. A more correct notion would be: “Influenza vaccination reduces the likelihood of suffering from disease and death caused by influenza virus infection.”
Funding

European Scientific Working Group on Influenza (ESWI), Belgium.

Conflict of interest

WEPB has held consultancies with pharmaceutical companies. JEM has received honoraria, consultancies, reimbursement for travel to meetings and research grants from the pharmaceutical industry. DJS reports no conflicts of interest. ASM has received consulting fees from pharmaceutical companies. JSNVT has received funding to attend influenza related meetings, lecture and consultancy fees and research funding from several influenza antiviral drug and vaccine manufacturers. Research funding from GlaxoSmithKline, AstraZeneca and F Hoffmann-La Roche is on-going; all forms of personal remuneration ceased in September 2010. He is a former employee of SmithKline Beecham plc. (now GlaxoSmithKline), Roche Products Ltd. (UK) and Aventis-Pasteur MSD (now Sanofi-Pasteur MSD), all prior to 2005, with no remaining pecuniary interests by way of share holdings, share options and accrued pension rights. A.D.M.E. Osterhaus is no longer CSO but a consultant to Viroclinics Biosciences BV, a contract research organization that collaborates with pharmaceutical companies and also CEO of Artemis BV, a organization which deals with infectious and non-infectious causes of wildlife disease.

Acknowledgements

The authors wish to thank Jan de Jong, Jos Nauta and Bram Palache for their assistance during all stages of this paper. DJS acknowledges European Union FP7 programs EMPERIE grant (223498), Human Frontier Science Program grant (P0050/2008), Wellcome grant (087982/A/IA), National Institutes of Health Director’s Pioneer Award (DP1-OD000490-01), and National Institute of Allergy and Infectious Diseases contract HHSN266200700010C.

Appendix A. Supplementary data

Supplementary material related to this article can be found in the online version, at http://dx.doi.org/10.1016/j.vaccine.2013.09.063.

References

Beyer, McElhaney, Smith, Monto, Nguyen-Van-Tam, Osterhaus:
Cochrane re-arranged: Support for policies to vaccinate elderly people against influenza.

Supporting Material 1: Biological and conceptual framework.

The terms and concepts used in our analysis are derived from the basic sequence of events in the patient journey: exposure, infection, outcome, and observation. Here, we summarize some important properties of these of four stages.

An exposure case is a person whose respiratory mucus layer has entrapped virus particles. The exposure risk \( R(E) \) in a population depends on properties of the virus (e.g., transmissibility) and the population (e.g., intensity of social contacts). Exposure is a 'silent' event, and \( R(E) \) cannot be assessed directly. In seasons without influenza virus circulation, \( R(E) \) is of course zero.

There are two possibilities for an exposed person: Either the intruding virus particles are eliminated, in whatever way, in the host's mucus layer (protection), or they succeed in passing this barrier, reaching suitable upper respiratory cells, and replicating there (infection). The infection risk \( R(I) \) depends on viral properties (e.g., particle dose, activity of viral neuraminidase) and host factors (e.g., thickness of mucus layer, presence of innate and specific immunity). In particular, specific neutralising antibody able to attach to virus particles in the mucus layer plays a major role in protection. The rationale of vaccination is to induce (or increase) this antibody. The time window between exposure and infection is the phase where the vaccine has to prove its ability to prevent infection (and all consequential infection-related events).

An infected person may show an outcome, or not. An outcome can be any consequence of infection: a specific immune response (e.g., antibody rise) with or without the generation of symptoms; clinical signs, symptoms and syndromes from typical self-limiting influenza-like illness (ILI) to serious non-fatal or fatal complications (e.g., pneumonia, secondary bacterial infection); and societal events (e.g., work absenteeism, hospitalisation). An outcome is characterised by the fraction of actual influenza infection cases, of all outcome cases, the infection fraction of the outcome. Ideally, it is 100%; then the outcome risk \( R(O) \) would directly reflect \( R(I) \). Outcomes with confirmation of virus replication by laboratory assays approach this ideal. When outcome is clinically or societally defined without confirmation of infection, outcome cases following influenza virus infection may not be distinguishable from those related to other conditions (mainly other co-circulating respiratory pathogens), and outcome does not only occur in persons infected by influenza but also in those non-infected
(Figure 1). Then, the infection fraction of the outcome is smaller than 100%, and $R(O)$ does not entirely reflect $R(I)$, but is exaggerated by the ‘noise’ from non-influenza conditions. On average, even without laboratory confirmation, the infection fraction is high for ILI, moderate for fatal and non-fatal complications, and (very) small for the outcome ‘all-cause death’.

**Figure 1: Decision tree for exposure, infection, and outcome.**

![Decision tree diagram]

$R_V(I)$, $R_U(I)$, $R_V(O)$, $R_U(O)$: infection risk and outcome risk in vaccinated and unvaccinated subjects, respectively.

The principal preventive effect in vaccinated people lies in the reduction of the infection risk $R(I)$ compared to those not vaccinated, because vaccine should induce or enhance antibody, and antibody should prevent infection after exposure. The infection risk reduction is usually expressed as $1 - R_V(I)/R_U(I)$, with $R_V(I)$ and $R_U(I)$ the infection risks in vaccinated and unvaccinated persons, respectively. Here, we call this measure **biological vaccine efficacy**, being aware of other definitions for the term ‘efficacy’ used in probability theory, epidemiology, and the influenza literature (see below).

In a season without influenza virus circulation (no exposure), biological vaccine efficacy is not defined, as $R(E) = 0$ and $R_U(I) = 0$, and therefore senseless; vaccine effects (risk reduction) can only occur when there is something to reduce. Biological vaccine efficacy occurs only in seasons with virus circulation.

Biological vaccine efficacy is directly related to the virus antigens and their antibody induction (e.g., vaccine formulation, dosage, lot) and follows their variation. Moreover, there is a certain seasonal variation, mainly related to antigenic drift of the influenza virus, resulting in a decreasing match between the intruding virus strain and host antibodies raised against the
vaccine strain. In case of antigenic shift, a vaccine can completely fail (i.e., biological vaccine efficacy = 0).

The \textbf{reduction of the outcome risk} \( R(O) \) by vaccination, within the virus circulation period, maybe called the \textbf{vaccine effectiveness against the outcome (VE)}: 1 minus \( R_V(O)/R_U(O) \), with \( R_V(O) \) and \( R_U(O) \) the outcome risk in vaccinated and unvaccinated persons, respectively. For an outcome entirely caused by influenza infection, and equal exposure risks in vaccinated and unvaccinated persons assumed, biological vaccine efficacy and VE would be of the same size. Outcomes with a lower infection fraction involve non-influenza cases, which cannot be prevented by influenza vaccination. Then, VE will be smaller than biological vaccine efficacy.

The decision tree in Figure 1 assumes that the vaccine (i.e., the vaccine-induced antibody) is efficacious only during the stage between exposure and infection, and that there is no additional antibody effect during the stage between infection and outcome. Thus, we do not consider here the case of a ‘leaky’ or infection-permissive vaccine [1], which does not (primarily) prevent infection but prevents its outcomes. We contend that this assumption is justifiable because, in the review period (1965 to 2005), mainly \textbf{parenteral inactivated influenza vaccine} was used. This vaccine type predominantly induces antibody against viral haemagglutinin (HA) and only traces of antibody against viral neuraminidase (NA). Anti-HA antibody protects by preventing infection; anti-NA antibody allows infection but protects by preventing outcome [2]. For future influenza vaccine types (e.g., T-cell vaccines), our assumption might not hold, and the concept would need to be revisited.

For a probabilistic context, Nauta et al. [3] have shown how VE depends, mathematically, on vaccine efficacy against infection (comparable to our biological vaccine efficacy), the occurrence of disease (outcome) in non-infected persons, and exposure risk. The latter two factors vary largely between years and geographic regions and are unrelated to vaccine efficacy, but greatly affect VE. Even when vaccine efficacy against infection is 100%, VE can reach values between 18% (sic) and 94%, given usual ranges for non vaccine-related factors.

This large variation is unavoidable, indeed inherent in VE measures and independent from \textbf{observation}. This latter stage concerns operational issues, i.e., where and how outcome cases are detected (study population, study design, etc.). Operational issues further contribute to variation of the VE measure, for example, through suboptimal vaccine storage or administration, or through imperfect case-finding strategies or outcome case definitions [4, 5]. Furthermore, when the study design is not randomized but observational, various sorts of \textbf{selection bias} can occur, which introduce differential vaccine uptake or unequal exposure risks in vaccinated and unvaccinated persons. These effects distort VE, which can even reach negative values (i.e.,
more outcome in vaccinees than in non-vaccinees). Recently, much discussion was carried on
about the so-called healthy user effect or frailty selection bias, this is preferential vaccination of
healthy persons with low risk of complications, while very frail elderly remain under-vaccinated
[6]. Selection strategies and adjustment techniques to reduce the impact of selection bias in
observational studies have been described and applied but were often not successful. Also
paradoxical mass treatment effects can occur: in highly vaccinated populations like those in
nursing homes, the exposure risk, and thus the assessed VE, may decrease while the
concerning population is highly protected.

Guided by the actual influenza pathology and the consequences for patients and health care,
we used only three broad outcome definitions:

- laboratory-confirmed disease,
- ILI, or any signs or symptoms, which indicate upper respiratory infection,
- non-fatal and fatal complications (morbidity and death from influenza-related causes).

The first alternative outcome definition is identical to Cochrane’s ‘influenza’ and concerns
persons with clinical signs and symptoms confirmed as influenza infection by laboratory tests.
The tests used in the original studies were virus culture from throat swab specimens, and
antibody assessment in paired sera (a ≥4-fold titre rise would indicate a positive case). Polymerase
chain reaction assays were not available until fairly recently. The resulting vaccine
effectiveness against laboratory-confirmed disease serves as closest available prediction
for biological vaccine efficacy.

The second outcome definition concerns persons who present with clinical signs and symptoms
indicating ILI in the broadest sense. ILI is generally self-limiting in children and non-elderly
adults and either stimulates consultation with a primary care professional or no medical
intervention at all, yet it causes school and work absenteeism. In the elderly, there is a risk of
permanent disability following acute influenza illness [7]. The resulting VE measure is the
vaccine effectiveness against ILI.

The third outcome definition generally concerns persons in whom the virus produces
complicated illness (e.g., pneumonia), or an exacerbation of underlying pulmonary, cardiac or
metabolic diseases. Particularly in the elderly, such complications are usually serious, deserve
secondary care intervention or hospitalisation, and can lead to death. The related VE measure
is the vaccine effectiveness against complications, a composite measure including different,
fatal and non-fatal, disease entities. The justification for this combination is biological and
clinical: all combined conditions have in common that they surpass the course of typical
influenza disease and necessitate secondary health care. Yet, these conditions have various infection fractions of outcome, which may introduce additional heterogeneity.

Our principal tool of analysis, simple frequency distributions of unweighted VE values, allows for heterogeneity. Frequency distributions can discriminate between different scenarios of virus circulation. Suppose a sufficiently large collection of trials and assume similar impacts of positive and negative accidental factors. Figure 2C shows the scenario of absence of virus circulation. Here, VE values are scattered at random between extremes, with a preference for extremes and zero. Their meta-analysed mean is zero. Predictions on biological vaccine efficacy cannot be drawn. This pattern would also occur if trials to assess the quality of a prophylactic malaria drug were performed in Alaska.

Figure 2: Expected frequency distributions of vaccine effectiveness values, for three basic scenarios.

<table>
<thead>
<tr>
<th>Virus circulation</th>
<th>No virus circulation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Graph" /></td>
<td><img src="image" alt="Graph" /></td>
</tr>
</tbody>
</table>

The maxima of the VE values in Figure 2A were around 22, 34, and 58%, derived from Table 1 in Nauta et al. [3] with 70% vaccine efficacy against infection, up to 10% exposure risk, 5% infection risk in unvaccinated persons, and RRD-values of 10 (VE against complications), 20 (VE against ILI), and 100 (VE against laboratory-confirmed disease), respectively. RRD, (relative risk of the disease for infected subjects versus non-infected subjects) is a measure for the infection-relatedness of the disease and corresponds with our infection fraction of the outcome.

The second scenario (Figure 2B) is that of vaccine failure: virus circulation but no vaccine efficacy (due to worthless vaccine formulations or low dosages, or to antigenic shift). Then the distributions of all three VE measures peak at 0%; their meta-analysed mean is again zero.

The first scenario (Figure 2A) considers virus circulation and a high biological vaccine efficacy (black area). Here, the entirety of VE values against laboratory-confirmed disease (green area) form a narrow distribution with a maximum close to, yet smaller than biological vaccine efficacy, depending on the laboratory assay. The distributions for VE against ILI (red area) and fatal and non-fatal complications (blue area) suffer from the diluting effect of co-pathogens, therefore they
are broader than the green distribution and their maxima show a larger distance to biological vaccine efficacy.

For completeness, all-cause death was also considered and analysed, although this outcome has a low infection fraction: on average, only 5% of all-cause deaths during winter season can be attributed to influenza [6, 8]. When assessed without bias, VE against all-cause death would be expected to peak between zero and 5%. However, observational studies assessing all-cause death, including those from our review period ending in 2005, often report extremely inflated estimates. A main bias in these studies may have been the healthy user effect.

Our conceptual framework may be objectionable as being too plain. We interpret the entirety of suitable VE values as an attempt to predict biological vaccine efficacy, the principal measure of interest. This interpretation explains the ease, with which we combined VE values from different settings, study designs, outcome definitions etc. The procedure can be refined and improved, if desired. For our main purpose, the test of whether there is a general biological vaccine efficacy in studies performed during four decades, this approach may be sufficient. The real data convincingly fit already with this plain framework.

In the influenza literature, several definitions for vaccine efficacy and effectiveness have been used. Proceeding from our biological framework, we propose another attempt to clarify these terms. Within the period from exposure to observed disease, the extra-cellular stage of virus stuck in the mucus layer of the upper respiratory tract is decisive for vaccine action. During this stage, vaccine-induced antibodies neutralise virus and thus prevent cell infection, or not. We linked this mode of vaccine action to the term biological vaccine efficacy as relative reduction of the infection risk. In this view, biological vaccine efficacy expresses the vaccine’s substantial ability to prevent infection. Vaccine efficacy simply ‘happens’ in a partly vaccinated population, observed or not. When we try to observe vaccine efficacy in clinical trials assessing outcome, we inevitably get to deal with additional accidental factors, environmental and operational, which variably blur the view on biological vaccine efficacy. We captured this situation by the term vaccine effectiveness (VE). Thus, while vaccine efficacy describes the molecular stage of antibody–virus interaction between exposure and infection, vaccine effectiveness describes what we see of the consequences when performing trials.

The Cochrane review, as with many other influenza publications, defines ‘vaccine efficacy’ as the reduction of the outcome ‘influenza’, which is laboratory-confirmed ILI. We call this measure ‘effectiveness against laboratory-confirmed disease’. To Cochrane’s definition of vaccine efficacy, Osterholm et al. [9], again in accordance with influenza literature, adds the requirement that ‘influenza’ is assessed in an RCT. Consequently, they define vaccine effectiveness as the
reduction of outcome assessed in observational studies. Of course, these definitions are established, yet we would object that here biological and clinical processes are mingled with operational issues. We think that it is conceptually purer to separate these domains. On the other hand, a part of this terminological discussion is rather academic. Osterholm et al. preferentially include trials with polymerase chain reaction assays for the detection of virus replication. The infection fraction of these assays tends to 100%, and therefore Osterholm’s measure very closely approaches our ‘biological vaccine efficacy’. Their result for adults (18 to 65 years), a meta-analysed mean of 59% (95%CI: 51 to 67%), is in surprising accordance with our main result for vaccine efficacy in the elderly: at least 50%, more likely ≥60%.

In summary, VE values from given trials depend on the biological vaccine efficacy, which is the principal measure for vaccine quality, but also on other, vaccine-unrelated factors. It is necessary to consider virus circulation (exposure level), vaccine-antigen properties and antigenic drift (infection level), non-influenza outcomes (outcome level), and study design and other operational issues (observation level). One or just a few trials cannot predict biological vaccine efficacy because of the variable nature of these other factors. For a meaningful prediction of the principal measure for vaccine quality, sufficiently large aggregations of individual VE values, and few but broad outcome definitions are required.

Literature


Supporting Material 2: Data Processing and Results.

Earlier versions of the Cochrane review on influenza vaccination in the elderly were published in 2005 [1] and 2006 [2]. We based our study on the latest update published in 2010 [3] and accepted verbatim the choices for inclusion and exclusion of publications made by the Cochrane authors. The Cochrane review retrieved 330 candidate publications, excluded 255 publications (77%) for reasons described, and presented detailed information about the 75 included publications ('Characteristics of included studies', pp. 33 to 88, in the following referred to as Cochrane Summaries). Numbers of outcome cases in vaccinated and unvaccinated subjects, the corresponding relative risk (RR) values or odds ratio (OR) values, and their confidence intervals (CIs), were given in 15 'comparisons' and 62 grand tables ('Data and Analyses', pp. 100 to 182, further on referred to as Cochrane Analyses) presenting 116 single strata of 248 VE values (Table 1). This compiled information was accepted for our own database without consulting the original publications; exceptions were indicated. When citing an original publication, we used Cochrane’s abbreviation (first author and year of publication, e.g., Ahmed 1995), usually without giving the complete reference, which can be found in the literature list of the Cochrane review (pp. 16 to 20).
Table 1: Sequence of analysis applied in the Cochrane review.

<table>
<thead>
<tr>
<th>Cochrane comparison</th>
<th>Influenza vaccine type</th>
<th>Study design</th>
<th>Population</th>
<th>Outcomes ('Grant Tables')</th>
<th>Antigenic drift / virus circulation (single strata)</th>
<th>Number of vaccine effectiveness values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>parenteral</td>
<td>cohort</td>
<td>nursing home</td>
<td>7</td>
<td>21</td>
<td>98</td>
</tr>
<tr>
<td>2</td>
<td>parenteral</td>
<td>cohort</td>
<td>community</td>
<td>10</td>
<td>24</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>parenteral</td>
<td>cohort</td>
<td>Community</td>
<td>8</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>parenteral</td>
<td>cohort</td>
<td>Community</td>
<td>8</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>parenteral + PCV²</td>
<td>cohort</td>
<td>community</td>
<td>4</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>adjuvanted parenteral</td>
<td>cohort</td>
<td>community</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>parenteral</td>
<td>adjusted cohort</td>
<td>community</td>
<td>5</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>parenteral</td>
<td>case control</td>
<td>community</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>parenteral + PCV</td>
<td>case control</td>
<td>community</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>parenteral + PCV</td>
<td>case control</td>
<td>nursing home</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>parenteral</td>
<td>adjusted case control</td>
<td>community</td>
<td>3</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>parenteral + PCV</td>
<td>adjusted case control</td>
<td>community</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>parenteral</td>
<td>RCT³</td>
<td>community / nursing home</td>
<td>4</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>inactivated aerosol</td>
<td>RCT</td>
<td>nursing home</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>live attenuated</td>
<td>RCT</td>
<td>nursing home</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td>62</td>
<td>116</td>
<td>248</td>
</tr>
</tbody>
</table>

1, not included: Comparison 16 (sensitivity analysis), Comparisons 17 and 18 (safety data, which were not examined here). ² PCV, pneumococcal vaccine. ³ RCT, randomised controlled trial.

We expressed a VE value as 1 minus RR value or 1 minus OR value, respectively. RR values from RCTs and unadjusted cohort studies, OR values from unadjusted case-control studies, and their variances were recalculated from the numbers of vaccinated and unvaccinated persons, as given in the Cochrane Analyses. Where zero number occurred in the denominator of a formula, it was replaced by 0.5, as proposed by Agresti [4]. For cohort and case-control trials, which were adjusted for several defined confounding factors, RR values or OR values and their standard errors as reported in the Cochrane Analyses were used verbatim. The RR and OR values as reported in the Cochrane Analyses and re-calculated here, were identical, with few disturbing exceptions. In ten cases with zero outcome number in vaccinated persons but non-zero in unvaccinated persons, Cochrane reported RR values between 0.06 and 0.60, while they are actually 0. In twelve cases with zero outcome number in unvaccinated persons, Cochrane reported RR values between 0.00 and 15.43 while they are actually not defined. Trials, in which the number of outcome cases in the unvaccinated arm is (very close to) zero, indicate a virus circulation too low, or an outcome too seldom, to produce measurable outcome cases. They result in actually meaningless VE values, which are prone to extreme random fluctuation. We
judged this to have occurred in trials with numbers of outcome cases less than five and decided to treat those trials as having taken place under periods of negligible virus circulation.

The relevant study characteristics varied considerably among the original publications, and large heterogeneity could be expected. We therefore preferred to present the data by drawing simple frequency distribution graphs, with a VE class width of 20% according to guidelines for frequency tables by Dawson-Saunders and Trapp [5], and allowing for negative values. We also meta-analysed the VE values and their variances to get an impression of the overall dispersion and to explore whether there was rough agreement with the corresponding frequency distributions. We applied the inverse variance-weighted method [6]. The weighting measure is the precision of the estimate (i.e., the inverse of the variance, favouring trials with large numbers and/or small variances). We chose this method instead of random-effect models [7] because VE values based on both RR and OR values had to be combined. As the calculation involves logarithmic transformation, a VE value of 1.00 (100% effectiveness) was replaced by 0.99.

Cochrane groups VE values according to 14 different outcome definitions, directly derived from the original publications (Table 2).

Table 2: Single outcome terms for vaccine effectiveness applied in the Cochrane review, and alternative grouping.

<table>
<thead>
<tr>
<th>Cochrane outcome definitions (in order of appearance in the Cochrane Analyses)</th>
<th>Number of vaccine effectiveness values</th>
<th>Alternative outcome definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILI (influenza-like illness)</td>
<td>40</td>
<td>ILI</td>
</tr>
<tr>
<td>Influenza (laboratory confirmed cases)</td>
<td>17</td>
<td>Laboratory confirmed disease</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>22</td>
<td>Complication (non-fatal)</td>
</tr>
<tr>
<td>Hospitalisation for ILI or pneumonia</td>
<td>12</td>
<td>Complication (non-fatal)</td>
</tr>
<tr>
<td>Death from influenza or pneumonia</td>
<td>32</td>
<td>Complication (fatal)</td>
</tr>
<tr>
<td>All deaths</td>
<td>26</td>
<td>All-cause death</td>
</tr>
<tr>
<td>Influenza cases (clinically defined without clear definition)</td>
<td>7</td>
<td>ILI</td>
</tr>
<tr>
<td>Hospitalisation for influenza or pneumonia</td>
<td>32</td>
<td>Complication (non-fatal)</td>
</tr>
<tr>
<td>Hospitalisation for any respiratory disease</td>
<td>29</td>
<td>Complication (non-fatal)</td>
</tr>
<tr>
<td>Deaths from respiratory disease</td>
<td>3</td>
<td>Complication (fatal)</td>
</tr>
<tr>
<td>Hospitalisation for heart disease</td>
<td>14</td>
<td>Complication (non-fatal)</td>
</tr>
<tr>
<td>Combined outcome: all deaths or severe respiratory illness</td>
<td>8</td>
<td>All-cause death</td>
</tr>
<tr>
<td>Hospitalisation for influenza or pneumonia or respiratory disease</td>
<td>5</td>
<td>Complication (non-fatal)</td>
</tr>
<tr>
<td>Pneumonia (no better defined)</td>
<td>1</td>
<td>Complication (non-fatal)</td>
</tr>
<tr>
<td>All</td>
<td>248</td>
<td></td>
</tr>
</tbody>
</table>
Meta-analyses with an abundance of outcome definitions may not be the optimal method to analyse these kinds of data. We chose to reduce the number of outcome definitions drastically, accepting a possible increase in heterogeneity within the alternative outcome classes. We used only three broad outcome definitions (Table 2, alternative outcome definitions): for laboratory-confirmed disease (N=17), ILI (N=47), and complications (N=150), plus the outcome all-cause death (N=34).

The available 248 VE values were further processed as follows (Table 3): VE values averaged over several seasons or combined for several trials were excluded, as VE value from one trial in one season was the unit of analysis. For most case-control trials and some cohort trials, Cochrane presented both unadjusted and adjusted values. To avoid data redundancy, only the adjusted ones were included; they were, by the way, virtually always larger than their unadjusted counterparts, which demonstrates a general adjustment effect. When a given trial assessed several Cochrane outcomes for non-fatal complications (e.g., 'hospitalisation for pneumonia and influenza' and 'hospitalisation for heart disease'), only the highest VE value was included in accordance with methodological advice on representation of a study with multiple outcomes by its highest effect size value, according to Devine [8]. In some trials, the study population was subdivided in people at risk for developing complications, and those not at risk; where possible, only VE-values covering the total population were included.

Then, we excluded trials with experimental vaccine types: an aerosol vaccine containing inactivated virus, and a live attenuated vaccine (one trial, respectively). In all remaining trials, established and marketed parenteral inactivated vaccines were used, with different valences (mono- to tetravalent, mostly bi- and trivalent). Whole virus, split or plain subunit formulations, but also adjuvanted subunit (one trial) or virosomal formulations (two trials), were used. Their ability to induce a sufficient antibody response in the elderly is similar [9]; therefore we treated them as one vaccine type. In a few trials, pneumococcal polysaccharide vaccine was co-administered with influenza vaccine; we disregarded this fact [10], but not the studies themselves. The antigenic content of vaccines used before 1978 was expressed in chick red blood cell agglutination (CCA) units; doses of 200 CCA units or higher were regarded as sufficient [11]. One trial performed in the 1967/68 season (d’Alessio 1969) used dosages of 150 CCA for the influenza A strains, which is regarded as suboptimal. Since 1978, antigenic content has been determined by the single radial immunodiffusion test [12] and expressed as μg HA; doses between 7 and 21 μg HA are regarded as sufficient [13]. As all trials since 1978 used commercially available vaccines within this dose range, we regard them as appropriate, except for one trial (Taylor 1992) [14], where a trivalent inactivated vaccine did not induce measurable
antibody\textsuperscript{1}; for reasons not identified. In total, eight VE values associated with experimental vaccine formulations or inadequate dosages were excluded.

Table 3: Inclusion and exclusion of vaccine effectiveness values.

<table>
<thead>
<tr>
<th>Description</th>
<th>Number of vaccine effectiveness values</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>All vaccine effectiveness values</td>
<td>248</td>
<td></td>
</tr>
<tr>
<td>Summary VE values covering several seasons</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Unadjusted VE values, where also adjusted ones are available</td>
<td>19</td>
<td>Excluded</td>
</tr>
<tr>
<td>(N=83)</td>
<td></td>
<td>(N=83)</td>
</tr>
<tr>
<td>More than one VE value for complications within one trial</td>
<td>20</td>
<td>Included</td>
</tr>
<tr>
<td>VE values for risk and no risk subgroups, where also those for total groups are available</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Experimental vaccine types, inappropriate doses</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>VE against laboratory confirmed disease</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>VE against ILI</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>VE against complications non-fatal</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>fatal</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>VE against all-cause death</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

Our final database consisted of 165 VE values, involving 2,504,162 elderly persons, presented in 95 trials and described in 64 publications. The smallest trial included 30 persons, the largest one 259,627. Trials took place in Europe, North-America, Asia, and Australia spanning four decades (from 1965 to 2005). Fifty-two trials were performed in community-dwellers and 43 in nursing home residents. Five trials were RCTs, while the vast majority of trials were observational (13 case-control trials and 77 cohort trials).

It is essential to identify trials performed during seasons without virus circulation (community), or without outbreaks (nursing homes). As outlined above, VE values from such trials cannot serve as predictors for biological vaccine efficacy. To identify these trials, we used the Cochrane Summaries, which contained semi-quantitative terms like ‘epidemic year’, ‘mild season’, ‘non epidemic’, or mentioned when the original publications did not provide information on virus

\textsuperscript{1} “Low acute hemagglutinating inhibition titers in vaccinated cases (comparable with the levels in unvaccinated cases) show that the vaccine was not immunogenic in the residents who became cases” (Taylor 1992, p. 96).
circulation at all. Five trials were indicated as ‘no epidemic year’ or ‘no outbreak’. For 35 VE values, the number of unvaccinated persons was smaller than five. We combined the two groups as outlined above. Figure 1 presents the resulting frequency distribution of these 40 VE values, with the expected pattern in the background (see Supporting Material 1, Figure 2C).

Figure 1: Distribution of VE values without virus circulation, or with small numbers of unvaccinated persons.

Number of VE values: 40. N_UO: Number of unvaccinated subjects with outcome. Expected pattern: see Figure 2C in Supporting Material 1.

The frequency distribution resembles, very closely, the expected pattern: VE values are scattered at random between extremes, with elevations at the extremes and zero; the meta-analysed mean is close to zero (+8.5%, 95% CI: -0.5 to +16.6%).

To explore how the VE values behave with decreasing virus circulation and how the large number of trials without information on virus circulation should be treated, we arranged the data according to four classes derived from the terms as given in the Cochrane Summaries: manifest virus circulation (epidemic or outbreak, including ‘mild epidemic’ and similar terms), ‘probable’ (doubtful) virus circulation, ‘non epidemic’ and ‘no information’ (Figure 2).
A plausible pattern emerges: There is a gradient from epidemic to non-epidemic seasons (red line). The distribution of the values without relevant information appears mainly to resemble those with virus circulation. We therefore decided to leave these uncertain VE values in the dataset.

The outcome ‘all-cause death’ was analysed separately. Given a biological vaccine efficacy of 60% and a infection fraction of 5%, VE values against all-cause death would be expected to peak close to zero (3%). Only one all-cause death values met this expectation (Lopez Hernandez 1994, VE = 2.7%). All other VE values were much larger, with a range between 24 to 75%, a maximum between 50 and 70%, and a seemingly inflated meta-analysed mean of 48% (95% CI: 47 to 50%) suggesting the presence of a healthy user bias. We excluded the ‘all-cause death’ values at this stage.
The remaining 112 VE values allowed the examination of the vaccine performance during virus circulation (Figure 3).

**Figure 3: Distribution of VE values according to alternative outcome definitions.**

The VE values against non-fatal and fatal complication behaved similarly (not shown) and could be combined. Then, the VE values against combined complications peaked between 30 and 50%, with a meta-analysed mean of 28% (95% CI: 26 to 30%). VE values against ILI had a broad maximum between 30 and 70%; their meta-analysed mean was plausibly larger (39%; 95% CI: 35 to 43%) than those against complications. VE values against laboratory-confirmed disease showed a narrow distribution, which peaked between 50 and 70% (49%; 95% CI: 33 to 62%). Visual examination of the three distributions and comparison with those expected let us predict the biological vaccine efficacy in this dataset as being at least 50%, but likely around 60% or larger. This pattern remains stable also with regard to a time trend within the data (see below).

For the classification of antigenic drift, the Cochrane Summaries recorded information about the antigenic match between vaccine components and strains actually circulating during season, using terms like 'match', 'probably matched', 'no match', 'unknown' or 'absent'. In many summaries, detailed information with taxonomic names of vaccine and circulating strains was
provided. In other summaries, the issue of antigenic match was not mentioned at all, either because the original publication did not report about it, or the Cochrane authors did not record it. Cochrane eventually applied the dichotomy match/mismatch. One should realise that antigenic drift is not well encapsulated by such a dichotomous handling, as it is a gradual process driven by molecular changes in the viral HA. It can better be quantified by the antigenic distance (AD) derived from antibody chessboard titrations with ferret sera [15]. In an attempt to determine the actual antigenic distances between vaccine and circulating strains for all trials, we relied on the information in the Cochrane Summaries. Missing information (e.g., taxonomic names of vaccine and/or circulating strains) was frequent. We first tried to ‘back-fill’ from the original publications, and otherwise imputed information in the following way: For studies, where no vaccine composition was reported, the WHO vaccine recommendations for that particular season and hemisphere were used. Where no circulating strains were reported, the predominant circulating strains for that season and region were used according to the annual recommendations and reports given in *The Weekly Epidemiological Record* [16]. As these external data may not always sufficiently reflect the actual situation within a given local trial, we chose three broad AD classes: High similarity (0 to 1.9), minor antigenic drift (2.0 to 3.9), major antigenic drift (≥4.0). We were aware for trials performed during the antigenic shift from A-H2N2 to A-H3N2 subgroup in 1968 or the re-appearance of A-H1N1 in 1977.

The vast majority of VE values was associated with high similarity (AD 0 to 1.9, N=97, meta-analysed mean: 39%, 95%CI: 38 to 41%). Twenty-four VE values with minor antigenic drift produced a clearly lower maximum of the frequency distribution (not shown) and meta-analysed mean: 29% (22 to 35%), a plausible finding demonstrating the expected impact of antigenic drift on VE. Major antigenic drift occurred for only four VE values and concerned the transitions A/Beijing/353/89(H3N2) to A/Beijing/32/92(H3N2) in the 1992/93 season (AD = 7.8) and A/Wuhan/359/95(H3N2) to A/Sydney/5/97(H3N2) in the 1997/98 season (AD = 4.7). Their meta-analysed mean is slightly but not significantly lower than that for minor antigenic drift: 27% (22 to 31%). Thus, although a gradual impact of antigenic drift on vaccine effectiveness can be demonstrated in the data, it is worth observing that even under such conditions, vaccine still provided considerable protection. The data do not capture antigenic shift.

On the level of observation and operational issues, we studied the frequency distributions of the three main study designs (Figure 4). The few RCTs produced a maximum of between 50 and 70% VE (meta-analysed mean 51%, 95% CI: 38 to 62%). The case control studies, most of them performed in the community and adjusted for confounders, showed a maximum between 30 and 50%, with a relatively narrow dispersion. The cohort studies were more broadly distributed, with an indistinct maximum between 10 and 50%.
We also searched for time trends within the review period. Figure 5 presents the VE values according to the year of trial performance. RCTs (green) were performed in the 1960s and 1990s, but not in the years between. Cohort studies (blue) were introduced in the early 1970s, and case control studies (red) in the early 1980s. Thus, for one decade there were studies with only one design.
The VE values of the observational studies are scattered within wide clouds, yet a simple time trend analysis reveals that VE tends to decrease within the review period, similarly for both case control and cohort studies (blue and red trend line, respectively). The average VE substantially fell from 52% in 1971 to 26% in 2004. This is a relevant finding, not mentioned in the Cochrane review. One explanation could be a decreasing vaccine formulation quality. The formulations of the inactivated parenteral vaccine type changed during the review period, indeed: In the 1970s and 1980s, the first generation formulation (whole virus vaccine) was replaced by the newer split and subunit vaccines; after 2000, adjuvanted subunit vaccine became available. There were also changes in the standard vaccine doses. But meta-analyses of antibody studies, which are not influenced by mass treatment effects, have not detected any relevant differences between these vaccine formulations in primed populations, in particular the elderly, or any time trends (review periods: 1975 to 1995 [17], and 1978 to 2009 [9], respectively). Another possibility might be the general increase in life expectancy during the review period, successively increasing the fraction of very old and frail people within elderly populations, which in turn could affect the VE measures. From a study by Simonsen et al. in the US elderly population [18], we conclude that this circumstance is not likely to produce such a strong impact on VE. Thirdly, a general decrease of epidemic intensity (thus, decrease of exposure risk) could
have occurred in some regions, as has been observed in The Netherlands. At last and possibly related: a successive reduction of virus circulation in the elderly population (and possibly in transmitter groups like health care workers), due to annually increasing vaccine coverage, hampering virus transmission and decreasing exposure risk. This beneficial mass treatment effect would have a paradoxically negative impact on the VE measure, which decreases with decreasing exposure risk, falsely suggesting a decline in vaccine benefits. During the review period, the vaccine coverage in the elderly increased sharply in most developed countries. In the United States, for example, vaccination coverage in the elderly rose from 16% in 1972 to 65% in 2000 [18].

We could estimate the vaccination coverage in the Cochrane data as in cohort studies (but not in RCTs and case control studies), the fraction of vaccinated persons reflects the vaccination coverage rate of the population in the year of study performance. For 48 cohort trials, coverage rate estimates could be calculated. Indeed, there was a strong time trend: the average coverage rate rose from 43% in 1971 to 64% in 2004 (grey trend line in Figure 5). Thus, the possibility that decreasing VE-values and increasing coverage rates are causally related indicating a paradoxical mass treatment effect, cannot be excluded. This important notion should be kept in mind when performing observational studies in highly vaccinated populations.

Another body of supporting evidence is available: Since the work of Hirota et al. [19], Nauta et al. [20] and Coudeville et al. [21, 22], biological vaccine efficacy can now be estimated from pre- and post-vaccination titres in antibody studies. In this approach, every individual antibody titre is quantitatively associated with an infection risk, using a reference protection curve based on data from 26 trials in adolescents and adults. This association may also hold for the elderly. When applied to a published antibody dataset of annual vaccine registration trials with 2,008 observations in the elderly [23], the Coudeville approach leads to a biological vaccine efficacy estimate strikingly in agreement with our result: 62%.

**Literature**


Supporting Material 3: Criticism.

We could show that treatment of the Cochrane data according to a biological and conceptual framework resulted in plausible patterns for virus circulation and antigenic drift, and the existence of biological vaccine efficacy in the elderly - *quod erat demonstrandum*. Here we want to explore in more detail why the Cochrane authors, though using the same data, have not reached similar results.

1. Cochrane’s classification of virus circulation

The terms to describe the degree of virus circulation as given in the Cochrane Summaries served as the basis for Cochrane's classification. Cochrane compressed these terms to the dichotomies ‘epidemic year’ / ‘non epidemic year’, and ‘outbreak’ / ‘no outbreak’, respectively, and applied this classification to the Cochrane Analyses. For simplicity, we condensed these terms into the dichotomy ‘epidemic’ / ‘non epidemic’. Figure 1 shows the graphic pattern of this classification for the 165 included VE values, regardless of outcome.
The frequency distribution for ‘epidemic’ appears plausible: it shows a maximum to the right (between 30 and 49%) and a meta-analysed mean of 36% VE. The pattern for ‘non epidemic’, however, is unexpected and implausible: it has an indistinct maximum between 10 and 90%, and its meta-analysed mean, which should be around zero, is even larger (40%) than that for the ‘epidemic’ distribution. Moreover, the number of ‘non epidemic’ VE values is very large (N=50, 30%). It would be expected if every third to fourth winter season saw no influenza virus circulating at all. The actual frequency of seasons with negligible virus circulation is, however, far smaller, e.g., 15% in Dutch sentinel surveillance data for the period 1970 to 2011.

A closer look at Cochrane’s classification procedure (Table 1) reveals inconsistencies between the Cochrane Summaries and the actual classification. E.g., the summary of the publication Nichol 1994a and Nichol 1994c clearly reported an epidemic season, but its VE values were classified as ‘non epidemic’ by Cochrane. Pregliasco 2002 assessed ILI and hospitalisation during ‘low viral circulation’, according to the summary, but, confusingly, Cochrane classified the ILI estimate as ‘epidemic’ and the hospitalisation estimate as ‘non epidemic’. More remarkably, Cochrane classified most (but not all) values from publications reporting mild seasons as ‘non epidemic’. Cochrane did not explain these decisions.
Table 1: Classification of virus circulation, according to Cochrane.

<table>
<thead>
<tr>
<th>Terms from publication summaries</th>
<th>Number of VE values</th>
<th>Cochrane’s classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>‘epidemic year’ / ‘outbreak’</td>
</tr>
<tr>
<td>‘epidemic year’ / ‘outbreak’</td>
<td>100</td>
<td>93</td>
</tr>
<tr>
<td>‘low epidemic levels’ / ‘(relatively) mild season’</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>‘probably epidemic’</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>‘no epidemic’</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>No information</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>All</td>
<td>165</td>
<td>115</td>
</tr>
</tbody>
</table>

Finally, most (but not all) original publications, which had not provided any relevant information on virus circulation, were classified as ‘non epidemic’. As we could show, however, this classification does not appear appropriate (Supporting Material 2, Figure 2). Likely, the authors of the publications in question performed their studies during virus circulation, but did not mention that. This omission is substantial but does not justify classifying the related values as ‘non epidemic’.

Thus, Cochrane’s classification on virus circulation is counter-intuitive and does not meaningfully separate epidemic from non-epidemic seasons. This may have contributed to Cochrane’s impression of ‘uninterpretable’ data.

2. Cochrane’s classification of antigenic drift

The Cochrane review classified the terms on antigenic match between vaccine components and circulating strains as given in the Cochrane Summaries, to the dichotomous groups ‘vaccine matching’ and ‘vaccine matching absent or unknown’, here, in short, ‘match’ and ‘no match/unknown’. Figure 2 presents the frequency distributions for this classification for the 125 VE values we also used to assess antigenic drift (Supporting Material 2).
The event 'no match/unknown' occurs frequently in this discrimination (N=30, 24%), much more often that would be expected from epidemiological data. *E.g.*, Palache *et al.* found a frequency of 'antigenic mismatch' of only 7% during the period of 1982 to 1996 [1]. Nevertheless, the graph shows two distributions, which differ sufficiently in their maximum and their meta-analysed means ('match': 39%, 'no match/unknown': 28%). Thus, Cochrane’s discrimination appears plausible – yet with the reservation that it is actually not the difference between ‘match’ and ‘no match’, which is expressed, but rather that between ‘high similarity’ and ‘minor antigenic drift’ (Supporting Material 2).

The Cochrane distribution of ‘no match/unknown’ exhibits a second maximum between 70 and 90%. This is a counter-intuitive finding for trials ‘without match’ and deserves more explanation. When examining Cochrane’s classification process we noted a number of inconsistencies. For example, in the Summaries, Cochrane recorded the vaccine strains of a trial with live attenuated vaccine performed in 1996/97 (*Rudenko 2001*) as A/Leningrad/134/17/57 and B/Ann Arbor/60/69 and classified them in the meta-analyses as a ‘match’. Would influenza strains from 1957 and 1969 be expected to match with wild strains from 1996? However, the recorded strains are the *master strains of the live vaccine*, not the strains, which deliver haemagglutinin
and neuraminidase (in that trial: A/Texas/36/91; A/Nanchang/933/95 and B/Harbin/7/94). The Cochrane authors appear to be unfamiliar with influenza vaccine production.

The study Currier 1988 reported A/Leningrad/360/86 as the vaccine strain, and an A/Leningrad-like strain as the circulating strain, but Cochrane classified the pertinent VE value as 'no match'. For Ruben 1974, Cochrane reported the vaccine strain as A/Aichi/2/62 and the circulating strain as A/England/42/72, and consequently classified the pertinent VE values as 'no match'. But the vaccine strain was A/Aichi/2/68 (H3N2), antigenically close to the A/Hong Kong/68 (H3N2) pandemic strain, and the circulating strain was the consecutive H3N2 variant exhibiting minor drift. Already the title of Ruben’s publication (“Effectiveness of killed Hong Kong vaccine against infection with the England strain”) would have given a hint, but the Cochrane authors appear to have overlooked this.

Table 2: Classification of antigenic match, according to Cochrane.

<table>
<thead>
<tr>
<th>Terms from Cochrane Summaries</th>
<th>VE values</th>
<th>Cochrane’s classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Meta-analysed mean (95% CI)</td>
</tr>
<tr>
<td>(probably) match</td>
<td>92</td>
<td>39% (38 to 40)</td>
</tr>
<tr>
<td>(probably) no match</td>
<td>14</td>
<td>28% (24 to 32)</td>
</tr>
<tr>
<td>unknown</td>
<td>13</td>
<td>47% (32 to 59)</td>
</tr>
<tr>
<td>not mentioned</td>
<td>6</td>
<td>15% (-5 to 32)</td>
</tr>
<tr>
<td>All</td>
<td>125</td>
<td>95</td>
</tr>
</tbody>
</table>

Table 2 shows how Cochrane translated the terms from the Summaries to the dichotomous match classes used in the Meta-analyses. The antigenic match of 19 VE values was not defined ('unknown' or not mentioned at all in the Summaries). Cochrane treated most (but not all) of these values as 'no match/unknown'. However, the VE values characterised as 'unknown' produce a frequency distribution (not shown) and a meta-analysed mean (47%, Table 2) larger than those with '(probably) match', suggesting that they also were associated with match. Indeed, the match of some of these values were 'unknown' only to the Cochrane authors, but can very well be defined using the available information. For example, the study Howells 1975a describes a trial from 1971/72 in the UK with a vaccine including the A/Hong Kong/68 (H3N2) pandemic strain (Cochrane wrote A2/HK/68). Cochrane characterised the match as 'unknown' and classified it as ‘no match’, although Hong Kong/68 still circulated in the UK in that season, which is actually a perfect match. In the study Cuneo Crovari 1980, both vaccine and circulating strains were B/Hong Kong/8/73, but according to Cochrane their match was 'unknown'.
Similar to Cochrane’s treatment of virus circulation, the discrimination between ‘match’ and ‘no match’ suffered from misclassification and could therefore not contribute to an understanding or interpretation of the data.

3. Cochrane’s claim on community dwellers versus nursing home residents

The Cochrane review rightfully points at some important differences between studies performed in community settings and in nursing homes: e.g., nursing homes residents are usually older and suffer from more morbidity than community dwellers. Moreover, influenza outbreaks in semi-closed communities either reach lower exposure rates than in the community if the virus is ‘kept out’, but once introduced, exposure rates become far higher than in the community. Although it is not totally clear which of their strata the Cochrane authors compared for this purpose, they suggest that the average VE in community dwellers is only of ‘modest’ size and smaller than in residents. Such a claim could have implications for vaccination policy and discourage vaccination of the elderly living in the community. If true, the claimed difference would be difficult to understand: Why should older and more diseased subjects respond better to vaccination, even given differences in exposure risk?

When drawing frequency distributions for the two populations using Cochrane’s data, we confirm a seemingly large difference between the two groups, indeed (Figure 3A): Community dwellers (green) show a broad frequency maximum between 10 and 50% and a meta-analysed mean of 27% (95% CI: 25 to 29%), and nursing home residents a maximum between 30 and 90% and an surprising mean of 41% (38 to 45%). Is this finding real?
We explored reasons for the huge difference between the two groups and found an influence of year of trial performance: Trials in community dwellers were performed from 1982 onwards (median: 1996), but in nursing home residents from 1965 onwards (median: 1982). Figure 5 in Supporting Material 2 shows that VE depends on year of performance: the older the trials, the larger, on average, the VE values. After adjustment for this imbalance using a simple meta-regression model (not shown), the difference in the crude data completely disappears (Figure 3B): now the two data subsets produce indistinguishable maxima (between 30 and 50%) and meta-analysed means (44 versus 42%). The claimed difference between community dwellers versus nursing home residents appears to be simply the result of a data imbalance with respect to the two observational study designs, and by a time trend towards lower VE values during the entire review period, two particularities in the data, which went unnoticed in the Cochrane’s analyses. We think that the factors which favour nursing home residents (higher exposure risk, on average) and community dwellers (lower age and morbidity, on average) balance in large numbers of VE values and that there might be no principal difference between the two sub-populations with respect to biological vaccine efficacy.

4. Data aggregation

The problem of over-stratification in the Cochrane review may be summarized by the fact that 248 identified VE values are arranged in 15 comparisons with 116 different strata (Table 1, Supporting Material 2), resulting in an average of only ~2 VE values per stratum. The Cochrane
authors have reported only one attempt to combine a number of strata (for assessing 'sensitivity', they mean the correlation between VE and 'study quality'): Comparison 16 (pp. 175 to 177) with 25 trials in 9211 persons for the outcome ILI produces a moderate but significant meta-analysed VE of 25% (13 to 35%). We regard this finding from a large number of trials with a similar outcome, as Cochrane’s most interesting result, yet it is not mentioned in the Abstract or the Plain Language Summary of the Cochrane review and will remain unnoticed by many readers.

Besides, we find Cochrane’s meta-analysed mean for VE against ILI underestimated. Our rearrangement produces a clearly higher value of 39% (35 to 43%). This discrepancy can be explained: Cochrane identified 40 trials with an outcome of ILI, but included only 25 in their comparison. Moreover, they included trials without virus circulation. But as already mentioned, nobody would accept a trial to assess a prophylactic malaria drug performed in Alaska, where the chance of getting exposed to malaria parasites is zero. We analysed such trials separately, confirmed that they behaved according to expectation, and proceeded with those trials performed during virus circulation. We believe that this procedure is more meaningful and more relevant to public health policy.

Literature