**S1. Mortality imputation**

We assumed that weekly R&C mortality was on overdispersed Poisson process, i.e.

 S1.1

for , where represents the index week, e.g. corresponds to the MMWR week 1762 (October 2 through October 8, 2005). The strictly positive factor follows a Gamma distribution with equal shape and rate parameter (), i.e.

 (Gamma()

Note that this Poisson-Gamma mixture model corresponds to a negative binomial model with.

 was modeled as

i.e. the sum of a “baseline mortality” and an influenza excess mortality component,

The baseline component was modeled as

and the influenza excess mortality component as

where , , A and represent the virologic indicators for influenza A(H1N1), A (H1N1)pdm09, A(H3N2) and B incidence, respectively. These indicators are derived in Supplement S2. Weekly R&C mortality was thus allowed to depend on influenza activity in the current week as well as one and two weeks prior, respectively. Goldstein et al. (reference 55 in main text) found a two-week lag of influenza mortality most important, but we wanted to allow for other lag times as well. We therefore assumed the effect of influenza circulation on mortality to be “immediate” (same week) as well as with one and two weeks delay.

Both mortality components were multiplied with “scaling factors” ( and , respectively) which allowed for a temporal trend in each, i.e.

and



**eFigure 1.**  Observed (blue dots) and fitted (posterior median) weekly R&C mortality (65+) values (red line) using the described model. The time period shown corresponds to August, 2005 through July, 2013. The weekly R&C mortality was imputed for the period from January, 2013 through July, 2014. Note that the fitted line represents the “systematic” component of the model (equation 3 in text) without the “fudge factor”, (expression S1.1).

**S2. Incidence proxy**

Our goal is to prove validity of an “incidence proxy” (IP) for influenza to be used for the estimation of excess outcomes such as mortality due to influenza. This incidence proxy is constructed from:

* Weekly data on the number of assumed “opportunities for testing” (hence referred to as “outcomes”), . An example of this is the influenza-like illness (ILI) surveillance system (outcome surveillance) which is based on reports of participating providers on the number of visits due to ILI.  might be normalized, e.g. by the number of providers reporting in a particular week. Alternative sources of this type of data might include the number of outpatient visit in a particular health care system and number of visits due to a a group of diagnoses assumed to potentially trigger testing for influenza.
* Weekly viral surveillance data, consisting of numbers of specimens tested, , and numbers resulting in a “positive” test, ; those two numbers are combined as in the “proportion positive” 

 represents a time (week) index such that  is the cumulative number of events  during week .  represents all outcomes, for example acute respiratory illness (ARI) in the population during week , observed and unobserved.  where  denotes a specific subset of , e.g. caused by influenza A(H3N2) and  is the complement of .  and  are phenomenologically identical, etiologically distinct: Caused by influenza virus, possibly specified by type/subtype *vs* caused by other agents, possibly by other influenza type/subtype viruses. Further, assume that the “observed” number or index of testing opportunities () is a random variable binomially distributed according to , where  is a constant “sampling ratio”. If  represents output of a surveillance system like the U.S. Outpatient Influenza-like Illness (ILI) Surveillance Network, the measure is strongly seasonally variable, not only because of seasonality of the outcome, but also because the numbers of providers reporting varies seasonally. As a result,  would no longer be constant and the following argument would fall apart. Therefore, the measure has to be normalized, e.g. by dividing by the number of providers reporting. In the reminder of the discussion we assume that  has been normalized. The test is of “perfect accuracy”: Every subject whose current health care visit precipitates testing is correctly classified as ill due to the agent of interest or not. The observed number of positive tests in week , , is driven by the prevalence of the outcome of interest in all cases at time ,  and by the sample  of all outcomes that are tested. Therefore, the number of positive tests  is distributed according to a hypergeometric distribution, i.e.

 

 The expected number of positive tests is given by

 

and the expected proportion positive is

 

Similarly, the expected number of observed outcomes is

 

Consider the proportion positive multiplied with the number of total outcomes observed, , as incidence proxy. For it to be “good”, we would require it to “track” the observed incidence of interest . By “tracking” we are referring to the the property of proportionality, i.e.:

 

where .

 and  are independent conditionally on ,  and . These three variables define epidemiological scenarios and can, in this context, be interpreted as parameters. Consider a given epidemiological situation, with a certain level of overall ARI incidence (), with a certain prevalence of the outcome of interest ()  and a given sampling ratio . The proportion testing positive for the cause of interest then only depends, besides the test characteristics, which for now we assume to be perfect, on the sample tested. That realize proportion positive  will, on average, be  (hypergeometric mean); and that clearly has nothing to do with the incidences of the outcome observed () and is thus stochastically independent of . Similarly, the number of outcomes observed  will, on average, be  (Binomial mean); again, that number should be independent on the proportion testing positive. However, scenarios are plausible, where that may not be true. For example, a high level of sample positivity might induce a higher will to report  etc. But here, by assumption, , the reporting ratio, is constant.

Therefore



 

 

Letting  we complete the proof and conclude that  is proportional to  and that, therefore,  is a valid proxy of .

**S3. Construction of sampling distributions**

To construct empirical 95% confidence intervals for the seasonal number of deaths averted by vaccination we first constructed sampling distributions for the age group-specific seasonal VE, monthly VC and monthly excess mortality.

For the seasonal VE (Table 1) and the monthly VC we used published point estimates and 95%-confidence intervals to construct Normal sampling distributions of the underlying maximum likelihood estimates (MLEs). To construct sampling distributions, we applied “natural” transformations to these estimates. The purpose of these transformations was to reconstruct the MLE giving rise to the respective published values. VE point estimates and confidence limits were transformed as

,

where corresponds to perfect VE. Estimates of cumulative VC were transformed as

,

where is expressed as a proportion. For each of these we reconstructed the associated standard error as , where and are the reconstructed upper and lower limits of the 95% confidence interval of , respectively and is the quantile of a Standard Normal Distribution such that . Pseudo-random numbers were then generated from the distributions and then back-transformed into the scale of the value of interest. This procedure is informally justified in Appendix 2. To construct sampling distributions for VE an additional step was required, because more than one estimate might be available for a particular season.

For the seasonal VE, we selected all estimates for VE against all influenza that were assessed in adults. These values were used to generate empirical VE samples for each imputed underlying MLE as described for the age groups under 5, 5-19 and 20-64.

We then generated 1,000 samples for each season, composed of samples from each constructed distribution. The number of samples contributed by each specific distribution was determined according to the formula

,

where .

Because of evidence of a decrease in vaccine effectiveness in older ages, but relative scarcity of direct results empirical VE sampling distributions were generated by a two stage process. If specific VE estimates existed for a specific season for 65+, which was the case only for the seasons 2005/6 onward, those estimates were used as described for the younger ages. If no specific estimates existed VE in that age group the empirical sampling distribution was generated by multiplying each value of the samples generated for the younger ages by a factor randomly selected from a Uniform distribution (0%, 60%). This corresponds to an average VE in the 65+ of 70% of that in the younger age groups and the uncertainty around that.

We generated an empirical distribution of monthly excess deaths by summing over MCMC samples indexed by the iteration number over the epidemiologic weeks belonging to particular months.

**S4. Averted deaths due to direct effect of vaccination as lower bound**

The method of estimating the number of deaths averted by vaccination captured by expression (1) in the main text assumes that vaccination only has a “direct” effect on mortality, ignoring the deaths that are avoided because of the resulting change in the force of transmission. Here, we provide an informal proof for the fact that the “true” number of deaths averted is necessily higher that the number estimated by using (1).

Let denote a homogeneously mixing population of size *N* in which deaths (or other events of the outcome of interest) are caused by influenza (or any other vaccine preventable disease) during time period *t*. Assume that a proportion of is protected by vaccination at time and therefore cannot suffer the outcome (death due to influenza). Therefore, a proportion of deaths will be averted by this “direct” effect of vaccination. Some of those effectively vaccinated, had they not been vaccinated, would have become infected with the possibility of infecting other people. Some of those people (secondary cases) might have perished because of influenza. Accodingly, the total number of deaths averted is and (expression 1 in main text) is a lower bound for the deaths averted.

**S5. Monte Carlo procedure**

We constructed Monte Carlo confidence intervals for the age group-specific numbers of deaths averted by vaccination by season [[34](#_ENREF_34)]:

1. For each of the 3,000 samples obtained from the MCMC procedure for monthly (age group-specific) excess mortality due to influenza, :
	1. Draw a value from the sampling distributions of the parameters used in expression (1), i.e. monthly vaccination coverage , seasonal vaccination effectiveness ; Construction of these sampling distributions is described in Supplement S3.
	2. Use these values to calculate .
2. Determine 2.5th percentile, median and 97.5th percentile of the resulting empirical distribution of to construct 95% empirical confidence intervals.

**S6. Sensitivity Analysis**

We investigated the sensitivity of our averted death estimates by variying the assumed VE from 10% to 70% and calculating averted death estimates using excess mortality and coverage estimates for the seasons 2011/12 and 2012/13.

|  |  |  |
| --- | --- | --- |
| **VE (%)** | **2011/12**† | **2012/13** |
| 10 | 750 (628,878) | 2129 (1785,2481) |
| 20 | 1621 (1358,1897) | 4589 (3846,5354) |
| 30 | 2646 (2216,3097) | 7468 (6254,8717) |
| 40 | 3870 (3239,4535) | 10887 (9103,12712) |
| 50 | 5354 (4487,6280) | 15010 (12544,17540) |
| 60 | 7196 (6023,8428) | 20077 (16765,23469) |
| 70 | 9545 (7978,11186) | 26463 (22113,30926) |
| † Median number averted (95% empirical confidence interval) |

**eTable S6.1** Estimated averted deaths in the ages 65+ yrs. during influenza seasons 2011/12 and 2012/13, with different levels of vaccine effectiveness.