identify the presence of severe illness or frailty.\textsuperscript{2} Including the diagnosis code variables in analytic models, therefore, does not produce estimates of effect that are less biased; in fact, this method appears to result in estimates of effect that are, if anything, more biased than the unadjusted estimates.

We agree that there are many published studies reporting a decreased risk of death and hospitalization in influenza vaccinated compared with unvaccinated seniors. Had we restricted our analyses to the influenza season, our findings could be similarly interpreted. However, reproducibility is not proof of validity. Our results clearly demonstrate that the differences in risk are not specific to the influenza season and so are biased. As we stated in our paper, we do not believe one can conclude that influenza immunization has no benefit with respect to mortality during the influenza season. Rather, we believe that because of the limitations of the data sources used, the non-randomized studies conducted to date do not provide a reliable basis for estimating the presence and/or size of this benefit.

---

**Benefits of influenza vaccine in US elderly—appreciating issues of confounding bias and precision**

**EELKO HAK,\textsuperscript{1*} ARNO W HOES,\textsuperscript{1} JIM NORDIN\textsuperscript{2} and KRISTIN L NICHOL\textsuperscript{3}**

Sir,

In a recent issue, Jackson \textit{et al.}\textsuperscript{1,2} raised the important issue of confounding in non-randomized studies and questioned the validity of effect estimates of influenza vaccination among elderly persons in such studies. Randomization may guarantee that groups are comparable with regard to observed and unobserved prognostic factors, and, therefore, randomized controlled trials (RCTs) are considered the paradigm to study vaccine effects.\textsuperscript{3} An important feature of randomization, if conducted properly, is that it removes all kinds of selection and confounding biases as a result of preferences. Clearly, results of large enough vaccine trials in which the primary outcome is clinically relevant (e.g. development of fatal or non-fatal disease) rather than a surrogate outcome (e.g. antibody response) provide crucial information on the true impact of vaccinations on health care and are best suited to prioritize such interventions.\textsuperscript{4} However, influenza vaccines are indicated for a wide variety of patients and clinically relevant outcomes are infrequent, thus a RCT on vaccine effects requires large representative samples. Importantly, vaccinations can only be effective when patients are actually exposed to the virus and match circulating strains. Further, placebo-controlled influenza vaccine trials in the elderly are considered unethical, since current immunization recommendations in many countries worldwide include such vaccinations in this high-risk population.

Non-randomized vaccine effectiveness studies are alternatives to RCTs and mainly include (variations on) cohort and case–control study designs.\textsuperscript{5,6} Such studies are attractive because computerized medical records are commonly used and large health care databases are widely available. Although there are many advantages of conducting such studies including applicability in different patient populations, timeliness, reduction of costs, and increased feasibility, it remains a challenge to ascertain validity.\textsuperscript{7} The main drawback of non-randomized studies is the absence of random assignment of the vaccine to study subjects.\textsuperscript{3,5,6} By definition, the selection of patients for vaccination is influenced by relevant patient characteristics, constituting the indication, that are also related to clinical outcome. Such confounding is often referred to as ‘confounding by indication’. The indication is likely to be related to stable patient characteristics (e.g. age, gender), time-varying prognosis (e.g. presence and severity of diseases, functional status), and health behaviour. Typically and similar to most primarily physician-initiated therapeutic and preventive interventions, the exposed group comprises patients with more severe disease or higher risk than the unexposed group.\textsuperscript{5,6} Crude, uncontrolled, estimates of the association between exposure to vaccines and outcome in such studies, therefore, lead to underestimation of vaccine effects. Conversely, if vaccine

---

\textsuperscript{1} Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands.

\textsuperscript{2} Health Partners Research Foundation, Minneapolis, MN, USA.

\textsuperscript{3} Minneapolis VA Medical Center, University of Minnesota, Minneapolis, MN 55417 USA.

\textsuperscript{*} Corresponding author. Julius Center for Health Sciences and Primary Care, PO Box 85500, 3508 GA, Utrecht, The Netherlands.

E-mail: e.hak@umcutrecht.nl

---

**References**


\textsuperscript{doi:10.1093/ije/dyi069}

Advance Access publication 17 April 2006
exposure is less dependent on physician indication and typically associated with good functional status or healthy lifestyle, the exposed group may comprise persons with a better prognosis than the control group, and such confounding by indication is usually referred to as ‘healthy user bias’.1,2 In this case, crude effect estimates will overestimate the true vaccine effects.

To avoid flawed conclusions from non-randomized studies, a particular challenge to the investigator is to prevent and adjust for ‘confounding by indication’ in the design of data collection and analysis, and, preferably, to quantify its potential magnitude.5,6,8 A prerequisite is that common potential confounders and their relative contribution to the direction and magnitude of confounding are largely known and clinically understood. Restriction of the study population to patients with a similar prognosis or matching of exposed and unexposed patients beforehand on levels of important confounding variables may be a powerful technique to reduce confounding. Further, statistical adjustments for prognostic differences between comparison groups must be applied correctly to enhance validity and precision of the estimates. Two most commonly applied powerful statistical methods are available to adjust for observed confounders, i.e. stratification or regression analysis and the propensity score method.6

In this light we need to appreciate the findings of Jackson et al.1,2 In a previous large HMO cohort study we showed that age, gender, previous health care use, and presence of diseases of the heart, lungs, and kidneys, cancer, dementia or stroke all significantly increase the risk of hospitalization for pneumonia or influenza (P&I) and all-cause mortality, and predictions of the presence of the outcome on the basis of the model were much improved over chance alone (area under the receiver operating curve estimates ranged from 0.73 to 0.83).9 Importantly, prevalence of these risk factors were more common in vaccinated than non-vaccinated persons, not only in our study but also in the studies by Jackson et al.,1,2 hence they are potential confounders in their data. As expected, adjustments by regression analysis for differences in prevalence of disease covariates led to movement of effect estimates further from the null compared with a regression model including age and gender only. Such findings make a strong case for the presence of typical ‘confounding by indication’ leading to underestimation of true influenza vaccine effects. In contrast, the case-control study by Jackson et al. also showed some higher prevalence of proxies for low functional status among non-vaccinated compared with vaccinated persons. Adjustments for these variables only in a partial regression model showed changes of the effect estimates towards the null and the authors concluded that ‘healthy user bias’ must have overestimated the true vaccine effect.6 Although the 95% confidence intervals of crude and functional status adjusted estimates were largely overlapping and precision was at stake, a complete model showing the combined effect of the confounding factors was not presented and is required to judge the direction and magnitude of the total observed confounding bias in the effect estimate. Clearly, the observed confounding bias given as the difference between crude and adjusted estimate by medical history was more pronounced and in the opposite direction than confounding by proxies of functional status (23% vs 20%). Thus, it can be expected that the fully adjusted estimate remains far from the null. A propensity score analysis, incorporating all predictors of exposure status should, therefore, complement such analyses to gain precision and to account for potential interactions.

Further, though Jackson et al. restricted their study population to community-dwelling elderly, excluding residents from nursing homes, they also should have excluded persons likely to die during the pre-influenza season. Such persons would not be considered reasonable candidates for vaccination, which by definition induces a ‘healthy user bias’. Similarly, persons hospitalized during the pre-influenza season may have delayed access to vaccination, hence unvaccinated hospital residents would be more likely to be re-hospitalized, again making such a pre-influenza season an invalid reference. Finally, influenza is around during the few months before the actual epidemic and effects, though less, can be expected during the pre-influenza season. In addition, though the authors state that the benefits of vaccination should be limited to the period of viral circulation, studies have suggested that influenza illness in the elderly may be associated with prolonged impact on impaired functional status and mortality extending several months beyond the illness. Thus, their findings of persistent but diminished benefit with regard to mortality during the pre-influenza and post-influenza season are biologically plausible and do not support the issue of a healthy user bias.

The current literature on the beneficial effects of influenza vaccination includes a full range of studies, from randomized serologically confirmed studies, to observational laboratory-confirmed studies and a wide range of cohort and case-control studies.10 Most individual studies and several meta-analyses show that influenza vaccination reduces morbidity and mortality related to influenza in seniors. We have demonstrated that after careful examining confounding bias using adequate statistical techniques, Jackson’s study would probably come to the same conclusion.

Acknowledgements
This work was funded by The Netherlands Scientific Organization (VENI No. 916.56.109).

Conflict of interest: Professor Nichol has received grants from Sanofi Aventis Pasteur.

References
We write to follow up on the editorial on the use of statistical
aggregates. We focus on the reaction, in it, to the letter from two of us (MPC and DS) in the same issue suggesting that the
‘longer (average) life expectancy of popes relative to artists’
reflects the lives characterized by social instability, high-risk
behaviours and geographical mobility (and thus infection risk)
of the artists.’ The Editorial presented evidence that the
‘assumption of better behaviour by the popes is perhaps
unjustified’. We have now looked behind the summary
behaviours were not—better behaviour ‘on average’ of popes is justified, these
life expectancy should be reversed. Even if the assumption of
longevity statistics, and present individualized data showing
that age at which each papacy commenced
traumatic test—another under-used analytic device—applied to Figure 1 reveals that among those who were alive at the
age at which each papacy commenced, the average remaining life of
the popes was shorter than that of the corresponding peer artists—at least up until 1750 or so, after which the
distributions became more similar.

The principal cause of this reversal is the phenomenon that the first analysis of this dataset sought to remove, namely that ‘Popes had to have reached a certain age before being elected to the papacy’. In that analysis, the statistical approach did not
fully address this constraint. Ideally, for each papacy-specific
‘longevity competition’, the time-clock should start when the
pope is elected, and the competition should include the pope,
and those artists born the same year as he, who were still alive
when he was elected. However, for several papacies, such
detailed matching is not possible. Instead, for each of the
1200–1599 papacies, the previous analysis effectively ‘started the clock’ at age 39—the age at which the youngest pope in that era was elected—by excluding artists who died before reaching that age. For the 1600–1900 papacies, it was started at age 38.

Unfortunately, under this broad scheme, as is clear from
Figure 1, several artists included in that analysis died before
‘their’ (and several other) pope(s) were even elected. This
inbuilt survival advantage for the popes is an example of what is today called ‘immortal time bias’. William Farr
described this fallacy in 1843. He noted that the average
age at death of bishops is greater than that of curates, and
thus—concerned for the underprivileged—suggested that
curates should be promoted to bishops, and at an early age,
‘for the sake of their health.’

Rather than match perfectly on year of birth and age at entry to
each longevity competition, one could for example proceed
half century by half-century, and determine the youngest age
\(A_{\text{min}}\) at which a pope born (or elected) in that half-century
was elected, and compare the post-\(A_{\text{min}}\) survival of these popes
and the corresponding artists. However, these half-century
(or even narrower) strata would still contain at least one other
pope elected at an age older than \(A_{\text{min}}\), after several artists
would already have died, and so the competition would
continue to be unfair.

In our new analysis, we circumvented this by creating a
separate contest (stratum) for each papacy. We started the
clock at the age at which the specific pope was elected. We
used as a comparison group those artists, born within 25 years
of when the pope was, who had reached that same age. For
example, in Figure 2, consider the papacy that began at 1335,
when the pope, born in 1280, was 55. Five ‘nearby’ artists,
born in 1260, 1266, 1280, 1284, and 1290, all of whom lived
until at least 55, serve as a comparison group. The pope died in
1342, at age 62, after 7 years as pope. His five ‘peers’ died in

---

6 Hak E, Verheij TJM, Grobbee DE, Nichol KL, Hoes AW. Confounding


9 Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events
is low and there are multiple confounders. Am J Epidemiol 2003; 158:280–87.


doi:10.1093/ije/dyl068
Advance Access publication 17 April 2006