Reporting Risk: From Math to Meaning

Lauren Lapointe-Shaw, MD PhD1–5, Glenda Babe, MA3, Peter C. Austin, PhD4,5, Andrew P. Costa, PhD4,6,7, Aaron Jones, PhD 4,6

1Division of General Internal Medicine and Geriatrics, University Health Network and Sinai Health System, Toronto, ON, Canada; 2Department of Medicine, University of Toronto, Toronto, ON, Canada; 3Women’s College Institute for Health System Solutions and Virtual Care, Women’s College Hospital, Toronto, ON, Canada; 4ICES, Toronto, ON, Canada; 5Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada; 6Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada; 7Centre for Integrated Care, St. Joseph’s Health System, Hamilton, ON, Canada

Corresponding Author: Lauren Lapointe-Shaw: lauren.lapointe.shaw@utoronto.ca

Submitted: 10 May 2022; Accepted: 18 July 2022; Published: 22 November 2022

DOI: https://doi.org/10.22374/cjgim.v17i4.638

Abstract
In this article, we discuss risk reporting and presentation in observational healthcare research. Reported measures of risk should facilitate clinical understanding. Although odds ratios are mathematically advantageous, they do not facilitate clinical understanding. Adjusted relative risk and risk difference are intuitive and clinically meaningful measures that can be readily obtained from any regression model using average marginal effects. Statistical summary measures or graphical displays that incorporate prevalence of risk factors can also be used to better identify the most important contributors to a target condition.

Résumé
Cet article se penche sur la communication et la présentation des risques dans le domaine de la recherche observationnelle sur les soins de santé. Les mesures du risque rapportées doivent faciliter la compréhension clinique. Bien que les rapports de cotes soient avantageux sur le plan mathématique, ils ne facilitent pas la compréhension clinique. Le risque relatif ajusté et la différence de risques sont des mesures intuitives et d’importance clinique, qui peuvent être facilement obtenues à partir de tout modèle de régression qui utilise des effets marginaux moyens. Les statistiques sommaires ou les représentations graphiques qui incorporent la prévalence des facteurs de risque peuvent également être utilisées pour mieux cerner les facteurs contributifs les plus importants à une maladie ciblée.

Keywords: measures of risk; risk difference; odds ratio; relative risk; average marginal effects; measurement
Introduction

Understanding risk is critical to clinical decision-making and essential for patient counseling. Consider the case of a primary primary care physician wondering which of their patients are at the greatest risk of being hospitalized with COVID-19. The results of a recent study\(^1\) demonstrated that individuals aged 75 years and older had 38 times the odds of being hospitalized than those aged 19–44 years. A health media reporter wonders whether to write that the risk is 38 times greater. The authors of the same paper received the following comment from a reviewer: “Please translate this result to something more meaningful to clinicians.” In another example, a public health physician considers piloting an intervention to keep borderline homebound people mobile outside their homes, and wishes to target this intervention to the most important risk factors. Should risk factors be prioritized based only on the strength of association? We start by discussing the limitations of different ways of reporting risk, and, through examples from previously published papers, demonstrate that more intuitive expressions of risk can be obtained using average marginal effects. We also examine how information on prevalence can be reported alongside the strength of association to provide a more fulsome picture of a risk factor’s contribution to population health.

Odds ratios—unintuitive yet mathematically advantageous

Physicians, like most other humans, tend to think in probabilities rather than in odds. This is apparent from how clinicians are taught to incorporate new knowledge into estimates of the probability of a diagnosis using likelihood ratios, which are odds-based. The calculation requires conversion between odds and probabilities, but clinicians and educators have instead been taught to use a conversion nomogram.\(^2\) These are impractical; as a result, once licensing examinations are passed, likelihood ratios and the odds that they represent are rarely considered again. Odds ratios provoke similar discomfort—only 19% of learners and 25% of speakers at an annual meeting of the Canadian Society of Internal Medicine (CSIM) understood odds ratios well enough to explain them to others.\(^3\)

Typically, the odds ratio (Table 1) has been interpreted as a relative risk (also known as a “risk ratio”) despite countless warnings not to do so.\(^4\) The clinician and health media expert in our introduction may have been tempted to express the study findings of Petrilli et al.\(^3\) as: “among those infected with COVID-19, those over age 75 had 38 times the risk of hospitalization as those aged 19–44.” Such a headline would surely attract a lot of attention. While odds ratios approximate relative risks when outcomes are rare, they overestimate them when outcomes are more common—this bias becomes important at outcome frequencies above 10%.\(^5\) For a given risk ratio, the odds ratio will always be more extreme (larger for odds ratios above 1.0, smaller for odds ratios below 1.0); however, the magnitude of bias would be highest with outcome frequencies above 10%, and for odds ratios that are farther from the null (farther from 1.0).\(^6\) In the study conducted by Petrilli et al., the overall rate of hospitalization (the outcome) was 52%, which is definitely not rare.\(^1\) The rate of hospitalization among those aged 75+ years was 93.7% compared to 23.7% among those aged 19–44 years. Although the authors do not report an adjusted relative risk, the unadjusted numbers suggest a relative risk not more than 4. Hence, a media headline stating a “38 times increase in risk of hospitalization” would have been off by nearly a factor of 10.

Fortunately, the results of randomized controlled trials are usually reported as risk differences and relative risks, facilitating interpretation. Yet, the majority of observational studies still report odds ratios. Why? Most observational studies use regression models to adjust for potential confounders, and the odds ratio has a key mathematical advantage in the regression modelling of binary outcomes.\(^4,6\)

Multivariable regression with a binary outcome models the probability that a patient experiences the outcome as a function of a set of covariates, which are typically risk factors of the outcome. The impact of each of these on the outcome probability could be expressed, theoretically, as a risk difference, relative risk, or odds ratio, depending on the type of regression model specified. Yet, regression models for risk differences and relative risks do not recognize that the outcome probability must fall between 0 and 1. This theoretical limitation is also practical, and an underlying reason why estimation methods for alternative models (e.g., linear probability, log binomial, and modified binary Poisson models) fail to converge or output impossible probabilities (i.e., above 1). For example, if a relative risk of 4 is applied to a baseline probability of 0.33 (33%), the result is 1.32 (132%)—yet the probability of an outcome cannot be more than 100%. Yet, if an odds ratio of 4 is applied to a baseline probability of 0.33 (33%, odds 1:2), the result is 0.66 (66%, odds 2:1). For this reason, some authors have argued that the odds ratio is the only “portable” measure of risk that can...
1.0 (100%) by the absolute risk difference,\(^6,13\) and are widely understood by clinicians and patients. The widespread adoption of NNT and NNH was a successful translation of absolute risk differences to a clinical audience—these have, fortunately, become a routine part of the interpretation of results from clinical trials. In contrast, results from regression-based observational studies are typically presented only in relative terms. Having understood that their older patients are only about four times (not 38 times) more likely to be hospitalized with COVID-19, a clinician may now wonder whether being of older age translates to a small, medium, or high risk of being hospitalized. Measures of relative risk are, by definition, numerically larger than absolute risk measures—and so without explicit reporting of absolute risk difference, clinicians may misinterpret the potential benefit (or harm) from a drug or treatment.\(^14–16\)

In regression-based approaches to confounder adjustment, measures of association are typically obtained directly from the logistic regression model rather than using an alternative model. Doi et al. recommend converting logistic regression outputs to relative risks and risk differences to aid clinical interpretation.\(^8\) Various conversion formulae have been developed\(^9–11\) to translate odds ratios to relative risk, but these have been criticized for overestimating the relative risk.\(^12\)

**Risk difference: a more intuitive expression of risk**

In randomized controlled trials, calculating risk difference is straightforward, using a 2 × 2 treatment × outcome contingency table. The numbers needed to treat (NNT) or numbers needed to harm (NNH) are easily derived by dividing 1.0 (100%) by the absolute risk difference,\(^6,13\) and are widely understood by clinicians and patients. The widespread adoption of NNT and NNH was a successful translation of absolute risk differences to a clinical audience—these have, fortunately, become a routine part of the interpretation of results from clinical trials. In contrast, results from regression-based observational studies are typically presented only in relative terms. Having understood that their older patients are only about four times (not 38 times) more likely to be hospitalized with COVID-19, a clinician may now wonder whether being of older age translates to a small, medium, or high risk of being hospitalized. Measures of relative risk are, by definition, numerically larger than absolute risk measures—and so without explicit reporting of absolute risk difference, clinicians may misinterpret the potential benefit (or harm) from a drug or treatment.\(^14–16\)

In regression-based approaches to confounder adjustment, measures of association are typically obtained directly

## Table 1. Review of common risk terms

<table>
<thead>
<tr>
<th>Definitions: Risk difference, relative risk, and odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary outcomes, such as 30-day readmission or in-hospital mortality, are common in health research. The epidemiological definition of the risk of an outcome is a simple probability, 0 to 1 (often expressed as a percentage). Both observational and experimental research frequently seek to quantify differences in the risk of outcomes between groups, typically defined by an intervention (experimental) or an exposure (observational) such as a socio-demographic, health status, or clinical history characteristic. For illustrative purposes, we will use the recent study conducted by Petrilli et al. on risk factors for the COVID-19 hospitalization,(^1) and specifically compare the risk of hospitalization between the oldest and the youngest age groups.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age group</th>
<th>Hospitalized N = 2741</th>
<th>Not hospitalized N = 2538</th>
<th>Total N = 5279</th>
</tr>
</thead>
<tbody>
<tr>
<td>Older (75+)</td>
<td>668 (A)</td>
<td>45 (B)</td>
<td>713</td>
</tr>
<tr>
<td>Younger (19–44)</td>
<td>437 (C)</td>
<td>1409 (D)</td>
<td>1846</td>
</tr>
<tr>
<td>Total</td>
<td>1105</td>
<td>1454</td>
<td>2559</td>
</tr>
</tbody>
</table>

In this study, 23.7% aged 19–44 years (437/1846) were admitted to hospital whereas this was 93.7% for those aged 75 years and older (668/713).

The simplest risk measure is the risk difference, which is obtained by subtracting one proportion from the other. This is termed an “absolute” measure, as it is expressed in the same units and scale as the outcome. In this case, the risk difference is calculated as 93.7% – 23.7% = 70.0%.

The relative risk (also known as “risk ratio”) is calculated by dividing the proportion with the outcome in one exposure group by the proportion with the outcome in the other exposure group: from our 2 × 2 contingency table, this is calculated as: \( \frac{A}{(A+B)} / \frac{C}{(C+D)} \). In this case, the relative risk is 93.7% ÷ 23.7% = 3.95 if comparing the higher risk (older age) group to the lower risk (younger age) group.

Finally, the odds ratio can be directly calculated from a 2 × 2 table, or from two proportions by first converting them to odds. To calculate the odds ratio from our 2 × 2 contingency table, the formula is \( \frac{A/B}{C/D} \). In this example, the odds ratio is (668/45) ÷ (437/1409) = 47.8. When outcomes are rare, \( A \) and \( C \) are small, and so \( \frac{A/B}{C/D} \) will approximate \( \frac{A/(A+B)}{C/(C+D)} \); that’s why the odds ratio approximates the relative risk when outcomes are rare. However, when they are common (as in this case), the odds ratio will be much larger than the relative risk. In this example, the odds ratio is (668/45) ÷ (437/1409) = 47.8.

Both relative risk and odds ratio are relative measures, meaning that they are not expressed on the same scale as the outcome but rather as a ratio.
Reporting Risk

from parameter estimates, which are the coefficients used in a regression formula. Yet, model parameter estimates do not directly translate to adjusted risk differences. This limitation is one of the reasons for the popularity of regression-free methods to minimize confounding in cohort studies. Propensity score weighting, as well as other forms of matching, can eliminate the need for multivariable regression. As a result, obtaining an easily interpretable relative risk, as well as an absolute risk difference from the matched (or weighted) groups, is straightforward and equivalent to how it would be done in an unadjusted analysis using a simple exposure by outcome $2 \times 2$ table.

As an alternative, what if instead of using parameter estimates to get risk measures from a logistic regression, we estimated adjusted probabilities of outcomes in each exposure group, and then used these to calculate measures of both relative and absolute risk? In the study by Petrilli et al., a logistic regression model was used to estimate the adjusted relationship between age and hospitalization for COVID-19. It is possible to estimate the adjusted probability of hospitalization in both younger and older age groups through a manual approach. The first step is to create two synthetic cohorts that are identical to the true cohort in all but one characteristic: in the first cohort, one sets every subject’s age to 19–44 years, and in the second set, every subject’s age is set to 75+ years. Then we estimate the risk of hospitalization for each subject in each of these fictitious cohorts using the logistic regression model specified using the full cohort. Next, we directly compare the average risk (proportion hospitalized) in the two cohorts. The difference between the two cohorts is the adjusted absolute risk difference for hospitalization attributable to being 75+ years age instead of 19–44 years age.

This quantity is termed the “average marginal effect” of 75+ years age, compared to 19–44 years age.

Petrilli et al. received a suggestion from a peer reviewer to present alternative adjusted measures of association that would be “more meaningful and useful to clinicians.” Thanks to this reviewer, the authors included average marginal effects, and we learned that those aged 75+ years had a risk of COVID-19 hospitalization that was an absolute 58% higher than in those aged 19–44 years. If this was a study of treatment, we would say in this case that the NNH was $100 \div 58\% = 1.72$ people.

Average marginal effects offer ease of estimation using existing regression models and effect measures on the scale of the outcome (e.g., count of days in a Poisson regression model for length of hospital stay, or as a proportion for a binary outcome, such as 30-day readmission). In addition to the manual approach we have described above, there are several macros available to facilitate the estimation of marginal effects in commonly used statistical packages: “margins” packages for SAS, R, and STATA output the average marginal effect, accompanied by 95% confidence intervals (CI). Whether the desired output is “marginal effects at the mean” (MEMs) or “average marginal effects” (AMEs) must be specified. MEMs are obtained by holding all other covariates at their mean values while manipulating a key exposure variable; however, these have been criticized for less realistically representing sample characteristics. In contrast, AMEs hold all other covariates at their natural (observed) values.

As AMEs manipulate one exposure while leaving all others at their observed values, they have been criticized for potentially reflecting unrealistic scenarios. Implausibility should not be a concern for exposures that could be randomized interventions applied to the entire sample or population. Yet, there are situations where manipulating only the main exposure variable could lead to implausible synthetic populations. An AME obtained from a regression model in which age is the main exposure and dementia is a covariate would be calculated as if the prevalence of dementia were the same across all age groups. As a result, there may be 20-year old persons with dementia in the synthetic cohort—an unlikely occurrence in the real world. However, similar concerns arise from regression coefficients whose values also represent the average effect across the entire sample, holding all other covariates fixed. For instance, when regressing an outcome on age (as the main exposure), dementia (as a covariate), and other baseline covariates, the effect of age can be interpreted as the effect of changing age while holding all other variables constant, including dementia.

Risk-reporting needs context

For a clinician, researcher, or policy leader seeking to improve the health of as many people as possible, the strength of association between a risk factor and an outcome only tells part of the story. Some strong risk factors may be relatively rare, and comparatively weaker risk factors, if common, may account for more cases. Although tools incorporating prevalence of a risk factor are used in classical epidemiology, this has not traditionally been the case in clinical epidemiology and healthcare research. The population attributable risk (PAR) incorporates the prevalence of a risk factor and its strength of association with the outcome, and can be estimated using an adjusted relative risk, adjusted risk
difference, or their components. As these values derive from the product of prevalence and risk difference, obtaining confidence intervals requires bootstrapping which can be computationally intensive.

Consequently, we propose a method of graphically representing risk factors in terms of both their prevalence and their strength of association in absolute terms. Prevalence of each risk factor can be easily obtained from the sample or the broader population, and an adjusted risk difference (with 95% CI) obtained using average marginal effects. A prevalence-difference plot provides a two-dimensional (2D) representation with intuitive interpretation: the upper left quadrant contains factors which are strongly associated with the outcome but less common, and the lower right quadrant contains factors that are more common but less strongly associated with the outcome. Factors in the lower left quadrant are both less common and more weakly associated with the outcome. Such a visual display translates the relative importance of different risk factors to a clinical or policy audience.

We illustrate this approach using data from a recent study of risk factors for homebound status among older adult homecare recipients in Ontario. We contrast information obtained from a traditional Forest plot of adjusted odds ratios with that obtained from a Forest plot of population-attributable risks (PARs), and our proposed prevalence-difference plot. In examining the forest plot of odds ratios (Figure 1), dependency for locomotion and winter season has the strongest associations with homebound status. However, Figures 2 and 3 provide more information by overlaying prevalence. In the forest plot of PARs (Figure 2), we have a reordering of factors with use of an assistive device and female gender now ranked highest. The prevalence-difference plot (Figure 3) affords an understanding of why this reordering occurred. While winter season and dependency for locomotion still have the strongest association with homebound status, and are associated with a 15–20% absolute increase in the risk of homebound, we also note that these risk factors are less than half as common as the use of an assistive device, which occurred in nearly 70% of the cohort. By considering only the first graph (Figure 1), we might conclude that locomotion dependency and winter season were more important drivers of homebound status than use of an assistive device, which would have been incorrect.

A 2D graphical display of risk factor prevalence and risk difference is a simple and intuitive method to demonstrate the relative contributions of different risk factors to overall cohort or population risk. It is less intensive computationally than the calculation of PARs with confidence intervals, and presents prevalence and risk difference separately, providing additional information on how each risk factor contributes to the overall risk. There are limitations to our approach. First, for measures of risk difference that are model-derived, their accuracy will depend on the appropriateness of the fitted model (in this example, logistic regression) and the adequacy of model fit. Second, depending on the sampling criteria, the prevalence and risk difference as measured in a cohort may not be reflective of their source population—any inference to the larger population must be made considering possible sources of sampling bias. These first two limitations would apply to all three methods of plotting risk. Third, although we believe that incorporating prevalence information adds considerably to the understanding of risk, we do not imply that this is the only contextual factor worthy of consideration when prioritizing risk factors for further research or policy change. It will also be essential to consider amenability and responsiveness of a particular risk factor to potential intervention.
Reporting risk in relative or proportional terms clouds meaning, and risk differences are more readily understood—these can be easily obtained from any regression using average marginal effects. Incorporating prevalence and other contextualizing variables further enhances meaning. We have presented several ways that risk reporting from regression-based observational studies can be improved to enhance clarity and aid clinical interpretation.

**Acknowledgments**

We thank Dr Christine Salahub for her assistance with data analysis, and Drs Peter Wu and Amol Verma for their helpful comments on an earlier version of this manuscript.

**Funding Statement**

The analysis in this study was supported by the Canadian Institutes of Health Research (grant number 148933) and ICES, which was funded by an annual grant from the Ontario Ministry of Health and Long-Term Care (MOHLTC). Dr. Austin was supported by a Mid-Career Investigator Award from the Heart and Stroke Foundation. Dr Lapointe-Shaw was supported by the University of Toronto, the Women’s College Institute for Health System Solutions and Virtual Care (WIHV), and The Peter Gilgan Centre for Women’s Cancers at Women’s College Hospital, in partnership with the Canadian Cancer Society.

**Author Contributions**

Lauren Lapointe-Shaw, Peter C. Austin, and Aaron Jones were responsible for conception and designing of the paper. Lauren Lapointe-Shaw, Glenda Babe, and Andrew P. Costa exercised acquisition, analysis, and interpretation of data for the paper. Lauren Lapointe-Shaw drafted the paper. All the authors revised the paper critically for important intellectual content, and approved its final version for publication.

**Declaration of Competing Interests**

None.

**Figure 2.** Forest plot of population-attributable risk (PAR) of homebound status from baseline characteristics. Using the summary data from a prior publication, we obtained PAR from the product of adjusted risk differences and risk factor prevalence in the cohort. 95% CI was obtained with 200 bootstrap replications. To facilitate comparison across Figures 1–3, we presented results for the same 10 characteristics in each graph.

**Conclusions**

Reporting risk in relative or proportional terms clouds meaning, and risk differences are more readily understood—these can be easily obtained from any regression using average marginal effects. Incorporating prevalence and other contextualizing variables further enhances meaning. We have presented several ways that risk reporting from regression-based observational studies can be improved to enhance clarity and aid clinical interpretation.
Figure 3. Plot of baseline characteristics according to adjusted difference in absolute risk of homebound and prevalence in a cohort. We term this a “prevalence-difference plot.” Characteristics in the top right represent those with a greater overall impact on the cohort-level risk of being homebound, as they are common and have a higher associated risk. Image reproduced from a prior publication.22
References