eAppendix for "On a square-root transformation of the odds ratio for a common outcome"

Let p_1 and p_0 denote the outcome probabilities with and without the exposure respectively, $RR = p_1/p_0$, and $OR = \frac{p_1(1-p_0)}{(1-p_1)p_0}$. The first proposition demonstrates that the square root transformation always moves the OR towards the RR.

Proposition 1. For causative exposures with $p_1 > p_0$, we have $OR - RR > sqrt(OR) - RR$; for preventive exposures with $p_1 < p_0$, we have $OR - RR < sqrt(OR) - RR$.

Proof. For causative exposures with $p_1 > p_0$, we have $OR = \frac{p_1(1-p_0)}{(1-p_1)p_0} > 1$ and thus $(OR - RR) - (sqrt(OR) - RR) = OR - sqrt(OR) = \frac{p_1(1-p_0)}{(1-p_1)p_0} - \sqrt{\frac{p_1(1-p_0)}{(1-p_1)p_0}} = \sqrt{\frac{p_1(1-p_0)}{(1-p_1)p_0}}\left(\sqrt{\frac{p_1(1-p_0)}{(1-p_1)p_0}} - 1\right) > 0$. If $p_1 < p_0$, we have $OR = \frac{p_1(1-p_0)}{(1-p_1)p_0} < 1$ and thus $(OR - RR) - (sqrt(OR) - RR) = \sqrt{\frac{p_1(1-p_0)}{(1-p_1)p_0}}\left(\sqrt{\frac{p_1(1-p_0)}{(1-p_1)p_0}} - 1\right) < 0$.

The next Proposition establishes bounds on the inflation that can occur for OR and $sqrt(OR)$ when the outcome probabilities themselves lie within certain bounds. That the bounds are sharp is established by the examples in the text.

Proposition 2. For causative exposures with $p_1 > p_0$, if $0.2 \le p_0, p_1 \le 0.8$, then $1 \le \frac{OR}{RR} \le 4$ and $\frac{4}{5} \le \frac{sqrt(OR)}{RR} \le \frac{5}{4}$; and if $0.1 \le p_0, p_1 \le 0.9$, then $1 \le \frac{OR}{RR} \le 9$ and $\frac{3}{5} \le \frac{sqrt(OR)}{RR} \le \frac{5}{3}$.

Proof. If $0.1 \le p_0, p_1 \le 0.9$, then $\frac{OR}{RR} = \frac{1-p_0}{1-p_1} \le \frac{1-0.1}{1-0.9} = 9$ and if $0.2 \le p_0, p_1 \le 0.8$ then $\frac{OR}{RR} = \frac{1-p_0}{1-p_1} \le \frac{1-0.2}{1-0.8} = 4$. We also have that $\frac{sqrt(OR)}{RR} = \sqrt{\frac{p_1(1-p_0)}{(1-p_1)p_0}}/\frac{p_1}{p_0} = \sqrt{\frac{p_0(1-p_0)}{(1-p_1)p_1}}$; if $0.1 \le p_0, p_1 \le 0.9$ then $\sqrt{\frac{p_0(1-p_0)}{(1-p_1)p_1}} \le \sqrt{\frac{0.5(1-0.5)}{(1-0.9)(0.9)}} = 5/3$ and $\sqrt{\frac{p_0(1-p_0)}{(1-p_1)p_1}} \ge \sqrt{\frac{0.1(1-0.1)}{(1-0.5)(0.5)}} = 3/5$; if $0.2 \le p_0, p_1 \le 0.8$ then $\sqrt{\frac{p_0(1-p_0)}{(1-p_1)p_1}} \le \sqrt{\frac{0.5(1-0.5)}{(1-0.8)(0.8)}} = 5/4$ and $\sqrt{\frac{p_0(1-p_0)}{(1-p_1)p_1}} \ge \sqrt{\frac{0.2(1-0.2)}{(1-0.5)(0.5)}} = 4/5$.

For preventive exposures with $p_1 < p_0$ by an analogous argument we have that if $0.2 \le p_0, p_1 \le 0.8$, then $\frac{1}{4} \le \frac{OR}{RR} \le 1$ and $\frac{4}{5} \le \frac{sqrt(OR)}{RR} \le \frac{5}{4}$; and if $0.1 \le p_0, p_1 \le 0.9$, then $\frac{1}{9} \le \frac{OR}{RR} \le 1$ and $\frac{3}{5} \le \frac{sqrt(OR)}{RR} \le \frac{5}{3}$. Proposition 3 establishes the relative proportion of values of p_1 and p_0 for which the inflation or deflation factor for $sqrt(OR)$, above to below the RR, is smaller than the inflation or deflation factor for OR, above or below the RR. By symmetry the proportion is the same for causative and preventive exposures.

Proposition 3. We have that

$$\begin{aligned}
\int_{p_0,p_1\in[0.1,1]} 1\left[\max(\frac{OR}{RR},\frac{RR}{OR}) > \max(\frac{sqrt(OR)}{RR},\frac{RR}{sqrt(OR)})\right] d(p_0,p_1) &\approx 0.93\\
\int_{p_1,p_0\in[0.2,1]} 1\left[\max(\frac{OR}{RR},\frac{RR}{OR}) > \max(\frac{sqrt(OR)}{RR},\frac{RR}{sqrt(OR)})\right] d(p_0,p_1) &\approx 0.99\\
\int_{p_1,p_0\in[0.25,1]} 1\left[\max(\frac{OR}{RR},\frac{RR}{OR}) > \max(\frac{sqrt(OR)}{RR},\frac{RR}{sqrt(OR)})\right] d(p_0,p_1) &= 1.
\end{aligned}$$

Proof. The factor by which the OR is inflated or deflated compared to the RR is given by $\max(\frac{OR}{RR},\frac{RR}{OR})$. The factor by which $sqrt(OR)$ is inflated or deflated, compared to the RR, is given by $\max(\frac{sqrt(OR)}{RR}, \frac{RR}{sqrt(OR)})$. To obtain the proportion of outcome probabilities for which the factor of inflation or deflation is greater for OR than for $sqrt(OR)$ we simply integrate the indicator function over the relative range of probabilities. The results in the Proposition are obtained by numerical integration.

The next Proposition establishes the bounds on the inflation or deflation of the OR and $sqrt(OR)$ as compared to the RR on the absolute scale. It can be seen from the proof that the bounds are sharp.

Proposition 4. For causative exposures with $p_1 > p_0$, if $0.2 \leq p_0, p_1 \leq 0.8$, then $|OR - RR| \leq 12$ and $|sqrt(OR) - RR| \leq 0.5509$; and if $0.1 \leq p_0, p_1 \leq 0.9$, then $|OR - RR| \leq 72$ and $|sqrt(OR) - RR| \leq 2.43$.

Proof. With $p_1 > p_0$, we have $|OR - RR| = OR - RR = \frac{p_1(1-p_0)}{(1-p_1)p_0} - \frac{p_1}{p_0} = \frac{p_1}{p_0}\left(\frac{(1-p_0)}{(1-p_1)} - 1\right)$ and if $0.1 \leq p_0, p_1 \leq 0.9$ then $\frac{p_1}{p_0}\left(\frac{(1-p_0)}{(1-p_1)} - 1\right) \leq \frac{0.9}{0.1}\left(\frac{(1-0.1)}{(1-0.9)} - 1\right) = 72$ and if $0.2 \leq p_0, p_1 \leq 0.8$ then $\frac{p_1}{p_0}\left(\frac{(1-p_0)}{(1-p_1)} - 1\right) \leq \frac{0.8}{0.2}\left(\frac{(1-0.2)}{(1-0.8)} - 1\right) = 12$. That $|sqrt(OR) - RR| \leq 2.43$ if $0.1 \leq p_0, p_1 \leq 0.9$, and that $|sqrt(OR) - RR| \leq 0.5509$ if $0.2 \leq p_0, p_1 \leq 0.8$ can be obtained by numerical grid search.

The next Proposition establishes the relative proportion of values of p_1 and p_0 for which the absolute differences $|OR - RR|$ are larger than $|sqrt(OR) - RR|$. Because differences in ratios are not symmetric, the results are reported separately for causative and preventive exposures.

Proposition 5. For causative exposures with $p_1 > p_0$ we have that

$$\int_{0.1 \leq p_0 \leq p_1 \leq 1} 1\{|OR - RR| \quad > \quad |sqrt(OR) - RR|\} d(p_0, p_1) \approx 0.95$$
$$\int_{0.2 \leq p_0 \leq p_1 \leq 1} 1\{|OR - RR| \quad > \quad |sqrt(OR) - RR|\} d(p_0, p_1) \approx 0.99$$

and for preventive exposures with $p_1 < p_0$ we have that

$$\int_{0.1 \leq p_1 \leq p_0 \leq 1} 1\{|OR - RR| \quad > \quad |sqrt(OR) - RR|\} d(p_0, p_1) \approx 0.90$$
$$\int_{0.2 \leq p_1 \leq p_0 \leq 1} 1\{|OR - RR| \quad > \quad |sqrt(OR) - RR|\} d(p_0, p_1) \approx 0.98.$$

Proof. To obtain the proportion of outcome probabilities for which the absolute difference $|OR - RR|$ is larger than $|sqrt(OR) - RR|$ we simply integrate the indicator function over the relative range of probabilities. The results in the Proposition are obtained by numerical integration.

Unfortunately, the same approach of taking a square root does not work as well with transforming hazard ratios (HR) to risk ratios, for outcomes that are common by the end of follow-up. Even if the outcome probabilities by the end of follow-up are between 0.2 and 0.8, a square root transformation of the hazard ratio can still be biased for the risk ratio by up to 48% (i.e. a factor of 1.48) and the untransformed hazard ratio by up to 80%; if the outcome probabilities by the end of follow-up are between 0.1 and 0.9, a square root transformation of the hazard ratio can still be biased for the risk ratio by up to 107% (i.e. a factor of up to 2.07) and the untransformed hazard ratio by up to 142%. However, a slightly different transformation to convert hazard ratios to risk ratios works well under a proportional hazard model. The transformation $RR \approx (1 - 0.5^{sqrt(HR)})/(1 - 0.5^{sqrt(1/HR)})$ reduces bias for the risk ratio considerably. If the outcome probabilities by the end of follow-up are between 0.2 and 0.8, this transformation is biased for the risk ratio by at most 16% (i.e. by a factor of at most 1.16), and if the outcome probabilities by the end of follow-up are between 0.1 and 0.9, this transformation is biased for the risk ratio by at most 45% (i.e. by a factor of at most 1.45). More formally, if we let $\lambda_1(t)$ and $\lambda_0(t)$ denote the hazard at time t for the exposed and unexposed respectively, we then have the following result.

Proposition 6. Under a proportional hazard model for the exposure such that $\frac{\lambda_1(t)}{\lambda_0(t)} = HR$ for all t, we have that

$$\begin{aligned}
\text{if } 0.2 &\leq p_0, p_1 \leq 0.8, \ \frac{1}{1.16} < \frac{(1 - 0.5^{sqrt(HR)})/(1 - 0.5^{sqrt(1/HR)})}{RR} < 1.16 \\
\text{and if } 0.1 &\leq p_0, p_1 \leq 0.9, \ \frac{1}{1.45} < \frac{(1 - 0.5^{sqrt(HR)})/(1 - 0.5^{sqrt(1/HR)})}{RR} < 1.45.
\end{aligned}$$

Proof. The survival function at the end of follow-up for the unexposed, $1 - p_0$, is given by $e^{-\int \lambda_0(t)d(t)}$ and the survival function at the end of follow-up for the exposed, $1 - p_1$, is given by $e^{-\int \lambda_1(t)d(t)} = e^{-HR \int \lambda_0(t)d(t)}$. We thus have the relation $(1 - p_1) = (1 - p_0)^{HR}$ and from this it follows $HR = \frac{\log(1-p_1)}{\log(1-p_0)}$. If we then use a numerical search for all values of the ratio

$$\frac{(1 - 0.5^{sqrt(HR)})/(1 - 0.5^{sqrt(1/HR)})}{RR} = \frac{(1 - 0.5^{sqrt(\frac{\log(1-p_1)}{\log(1-p_0)})})/(1 - 0.5^{sqrt(1/\frac{\log(1-p_1)}{\log(1-p_0)})})}{p_1/p_0},$$

when $0.2 \leq p_0, p_1 \leq 0.8$, there are no values that are above 1.16 or below 1/1.16; and when $0.1 \leq p_0, p_1 \leq 0.9$, there are no values that are above 1.45 or below 1/1.45.