Riley et al. (2021) (https://onlinelibrary.wiley.com/doi/full/10.1002/sim.9025) present in their **Equation 11** a formula for estimating the standard error of the C-statistic (SE(C)) as a function of the C-statistic, N (sample size) and ϕ (the estimated outcome proportion).

This equation is also presented in Fig. 3 of Riley et al (2024) (https://www.bmj.com/content/384/bmj-2023-074821).

Ideally, one would like to rearrange Equation 11 to obtain N as a function of the other parameters, thus allowing the calculation of the sample size required for the external validation of a model. The authors state that rearrangement of Equation 11 is not possible, and therefore estimate N via iterative approaches. The iterative approach is implemented in their software.

However Equation 11 can be rearranged, as shown below.

Rearrangement of Equation 11

Note: below we use the symbol Ns for N and Cs for C, as **N and C are protected symbols**. We also use SE2 to represent SE$(C)^2$.

Here is the representation of equation 11, in terms of SE$(C)^2$ to avoid dealing with the square root.

In[1]:= **Clear[SE2, Cs, Ns, ϕ];**

$$\text{SE2} = \frac{\text{Cs}\ (1-\text{Cs})\left(1+\left(\frac{\text{Ns}}{2}-1\right)\left(\frac{1-\text{Cs}}{2-\text{Cs}}\right)+\frac{\left(\frac{\text{Ns}}{2}-1\right)\text{Cs}}{1+\text{Cs}}\right)}{\text{Ns}^2\ \phi\ (1-\phi)}$$

Out[2]=
$$\frac{(1-\text{Cs})\ \text{Cs}\left(1+\frac{(1-\text{Cs})\left(-1+\frac{\text{Ns}}{2}\right)}{2-\text{Cs}}+\frac{\text{Cs}\left(-1+\frac{\text{Ns}}{2}\right)}{1+\text{Cs}}\right)}{\text{Ns}^2\ (1-\phi)\ \phi}$$

It is possible to represent SE$(C)^2$ in a simpler form:

In[3]:= **ExpandAll[SE2] // FullSimplify**

Out[3]=
$$\frac{(-1+\text{Cs})\ \text{Cs}\ (-2+2\ (-1+\text{Cs})\ \text{Cs}\ (-1+\text{Ns})-\text{Ns})}{2\ (-2+\text{Cs})\ (1+\text{Cs})\ \text{Ns}^2\ (-1+\phi)\ \phi}$$

Now make the substitution z = Cs-1:

In[4]:= **Clear[z];**

SE2 = ((-1 + Cs) Cs (-2 + 2 (-1 + Cs) Cs (-1 + Ns) - Ns)) /
(2 (-2 + Cs) (1 + Cs) Ns^2 (-1 + ϕ) ϕ) /. (-1 + Cs) → z

Out[4]=
$$\frac{\text{Cs}\ z\ (-2-\text{Ns}+2\ \text{Cs}\ (-1+\text{Ns})\ z)}{2\ (-2+\text{Cs})\ (1+\text{Cs})\ \text{Ns}^2\ (-1+\phi)\ \phi}$$

Further substitute `2 Cs - 2 = 2 z` and `Cs = z + 1`:

In[5]:=
```
SE2 /. {-2 + 2 Cs → 2 z, Cs → z + 1} // FullSimplify
```

Out[5]=
$$\frac{z\,(1+z)\,(-2-Ns+2\,(-1+Ns)\,z\,(1+z))}{2\,Ns^2\,(-1+z)\,(2+z)\,(-1+\phi)\,\phi}$$

We now have an expression for SE2 that is more amenable to algebraic manipulation. Here it is in input form:

In[6]:=
```
InputForm[%]
```

Out[6]//InputForm=
```
(z*(1 + z)*(-2 - Ns + 2*(-1 + Ns)*z*(1 + z)))/(2*Ns^2*(-1 + z)*(2 + z)*(-1 + ϕ)*ϕ)
```

Now solve for Ns in the expression:

In[7]:=
```
Clear[SE2, z, Ns, ϕ, sol];
sol = Solve[SE2 ==
   (z (1 + z) (-2 - Ns + 2 (-1 + Ns) z (1 + z))) / (2 Ns^2 (-1 + z) (2 + z) (-1 + ϕ) ϕ), Ns]
```

Out[8]=
$$\Big\{\Big\{Ns \to \Big(-z + z^2 + 4 z^3 + 2 z^4 - \sqrt{\Big(\big(z - z^2 - 4 z^3 - 2 z^4\big)^2 - 4\,\big(2 z + 4 z^2 + 4 z^3 + 2 z^4\big)\,\big(4\,SE2\,\phi - 2\,SE2\,z\,\phi - 2\,SE2\,z^2\,\phi - 4\,SE2\,\phi^2 + 2\,SE2\,z\,\phi^2 + 2\,SE2\,z^2\,\phi^2\big)\Big)}\Big) \Big/ \Big(2\,\big(4\,SE2\,\phi - 2\,SE2\,z\,\phi - 2\,SE2\,z^2\,\phi - 4\,SE2\,\phi^2 + 2\,SE2\,z\,\phi^2 + 2\,SE2\,z^2\,\phi^2\big)\Big)\Big\},$$
$$\Big\{Ns \to \Big(-z + z^2 + 4 z^3 + 2 z^4 + \sqrt{\Big(\big(z - z^2 - 4 z^3 - 2 z^4\big)^2 - 4\,\big(2 z + 4 z^2 + 4 z^3 + 2 z^4\big)\,\big(4\,SE2\,\phi - 2\,SE2\,z\,\phi - 2\,SE2\,z^2\,\phi - 4\,SE2\,\phi^2 + 2\,SE2\,z\,\phi^2 + 2\,SE2\,z^2\,\phi^2\big)\Big)}\Big) \Big/ \Big(2\,\big(4\,SE2\,\phi - 2\,SE2\,z\,\phi - 2\,SE2\,z^2\,\phi - 4\,SE2\,\phi^2 + 2\,SE2\,z\,\phi^2 + 2\,SE2\,z^2\,\phi^2\big)\Big)\Big\}\Big\}$$

There are two solutions provided. Resubstitute z = Cs -1 and simplify the expressions:

In[9]:=
```
sol = sol /. z → Cs - 1 // FullSimplify
```

Out[9]=
$$\Big\{\Big\{Ns \to \Big(Cs + Cs^2 - 4\,Cs^3 + 2\,Cs^4 - \sqrt{\Big((-1+Cs)\,Cs\,\Big((-1+Cs)\,Cs\,(1-2\,(-1+Cs)\,Cs)^2 + 16\,\big(-2+Cs-2\,Cs^3+Cs^4\big)\,SE2\,\phi - 16\,\big(-2+Cs-2\,Cs^3+Cs^4\big)\,SE2\,\phi^2\Big)\Big)}\Big) \Big/ \big(4\,(-2+Cs)\,(1+Cs)\,SE2\,(-1+\phi)\,\phi\big)\Big\}, \Big\{Ns \to \Big(Cs + Cs^2 - 4\,Cs^3 + 2\,Cs^4 + \sqrt{\Big((-1+Cs)\,Cs\,\Big((-1+Cs)\,Cs\,(1-2\,(-1+Cs)\,Cs)^2 + 16\,\big(-2+Cs-2\,Cs^3+Cs^4\big)\,SE2\,\phi - 16\,\big(-2+Cs-2\,Cs^3+Cs^4\big)\,SE2\,\phi^2\Big)\Big)}\Big) \Big/ \big(4\,(-2+Cs)\,(1+Cs)\,SE2\,(-1+\phi)\,\phi\big)\Big\}\Big\}$$

Numerically examine both solutions. Obviously Ns has to be positive.

In[10]:=
```
sol1 = Ns /. sol[[1]]
```

Out[10]=
$$\Big(Cs + Cs^2 - 4\,Cs^3 + 2\,Cs^4 - \sqrt{\Big((-1+Cs)\,Cs\,\Big((-1+Cs)\,Cs\,(1-2\,(-1+Cs)\,Cs)^2 + 16\,\big(-2+Cs-2\,Cs^3+Cs^4\big)\,SE2\,\phi - 16\,\big(-2+Cs-2\,Cs^3+Cs^4\big)\,SE2\,\phi^2\Big)\Big)}\Big) \Big/ \big(4\,(-2+Cs)\,(1+Cs)\,SE2\,(-1+\phi)\,\phi\big)$$

NOTE: in the Riley et al. (2021) paper, they report use of SE = 0.0255, but the Stata simulation code in the Supplementary material (and on which their presented results are based) uses 0.02551. Substitute for Cs, ϕ and SE2 and examine the numerical result (negative in this case):

In[11]:=
```
sol1 /. {Cs → 0.7, ϕ → 0.1, SE2 → 0.02551^2}
```
Out[11]= -1.1116

Now examine the 2nd solution. After substituting for Cs, ϕ and SE2 the numerical result is positive (so this is the expression we want):

In[12]:=
```
sol2 = Ns /. sol[[2]]
sol2 /. {Cs → 0.7, ϕ → 0.1, SE2 → 0.02551^2}
```
Out[12]=

$$\left(Cs + Cs^2 - 4\,Cs^3 + 2\,Cs^4 + \sqrt{\left((-1+Cs)\,Cs\,\left((-1+Cs)\,Cs\,(1-2\,(-1+Cs)\,Cs)^2 + 16\left(-2+Cs-2\,Cs^3+Cs^4\right)SE2\,\phi - 16\left(-2+Cs-2\,Cs^3+Cs^4\right)SE2\,\phi^2\right)\right)}\right) \Big/ \left(4\,(-2+Cs)\,(1+Cs)\,SE2\,(-1+\phi)\,\phi\right)$$

Out[13]= 1153.03

sol2 in input form:

In[14]:=
```
InputForm[sol2]
```
Out[14]//InputForm=
```
(Cs + Cs^2 - 4*Cs^3 + 2*Cs^4 + Sqrt[(-1 + Cs)*Cs*((-1 + Cs)*Cs*(1 - 2*(-1 + Cs)*Cs)^2 +
     16*(-2 + Cs - 2*Cs^3 + Cs^4)*SE2*ϕ - 16*(-2 + Cs - 2*Cs^3 + Cs^4)*SE2*ϕ^2)])/
 (4*(-2 + Cs)*(1 + Cs)*SE2*(-1 + ϕ)*ϕ)
```

3.3.1 Illustrative example

The authors (Riley et al. 2021) use two sets of inputs to illustrate the calculation of Ns.

First set of inputs: Cs = 0.7, ϕ = 0.1, SE = 0.02551 (i.e., SE$(C)^2$ = 0.02551^2). The reported Ns was 1154. Use the Ceiling function to round up:

In[15]:=
```
sol2 /. {Cs → 0.7, ϕ → 0.1, SE2 → 0.02551^2}
% // Ceiling
```
Out[15]= 1153.03

Out[16]= 1154

Second set of inputs: Cs = 0.8, ϕ = 0.5, SE = 0.02551. The reported Ns was 302.

In[17]:=
```
sol2 /. {Cs → 0.8, ϕ → 0.5, SE2 → 0.02551^2}
% // Ceiling
```
Out[17]= 301.771

Out[18]= 302

4.3 Step (iii): Sample size for C-statistic

Inputs: Cs = 0.8, ϕ = 0.018, SE = 0.02551. The reported Ns was 4252.

In[19]:=
```
sol2 /. {Cs → 0.8, ϕ → 0.018, SE2 → 0.02551^2}
% // Ceiling
```
Out[19]= 4251.43

Out[20]= 4252

As a sensitivity analysis, they also considered C - statistics of 0.75 and 0.85, which suggested sample sizes of 5125 and 3271 participants, respectively.

In[21]:=
```
sol2 /. {Cs → 0.75, ϕ → 0.018, SE2 → 0.02551^2}
% // Ceiling
```
Out[21]= 5124.14

Out[22]= 5125

In[23]:=
```
sol2 /. {Cs → 0.85, ϕ → 0.018, SE2 → 0.02551^2}
% // Ceiling
```
Out[23]= 3270.65

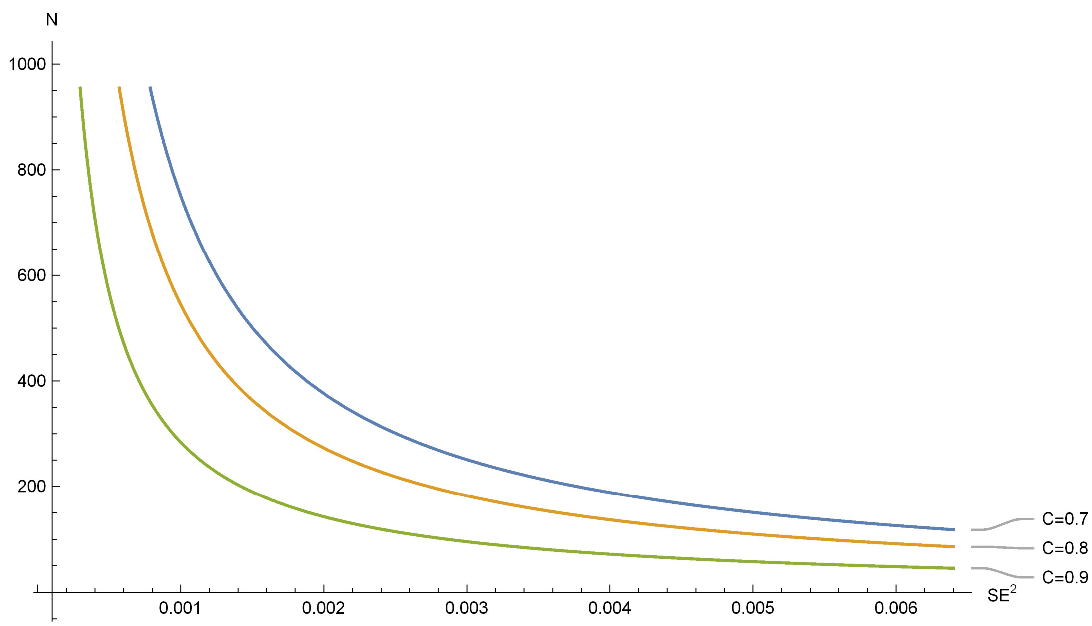
Out[24]= 3271

Plotting the solution

Estimated sample size *N* as a function of SE(*C*)^2 for different C-statistic values.

In[25]:=
```
c1 = sol2 /. {Cs → 0.7, ϕ → 0.1};
c2 = sol2 /. {Cs → 0.8, ϕ → 0.1};
c3 = sol2 /. {Cs → 0.9, ϕ → 0.1};
Plot[{c1, c2, c3}, {SE2, 0.01^2, 0.08^2},
 PlotLabels → {"C=0.7", "C=0.8", "C=0.9"}, AxesLabel → {"SE^2", "N"}, ImageSize → Large]
```

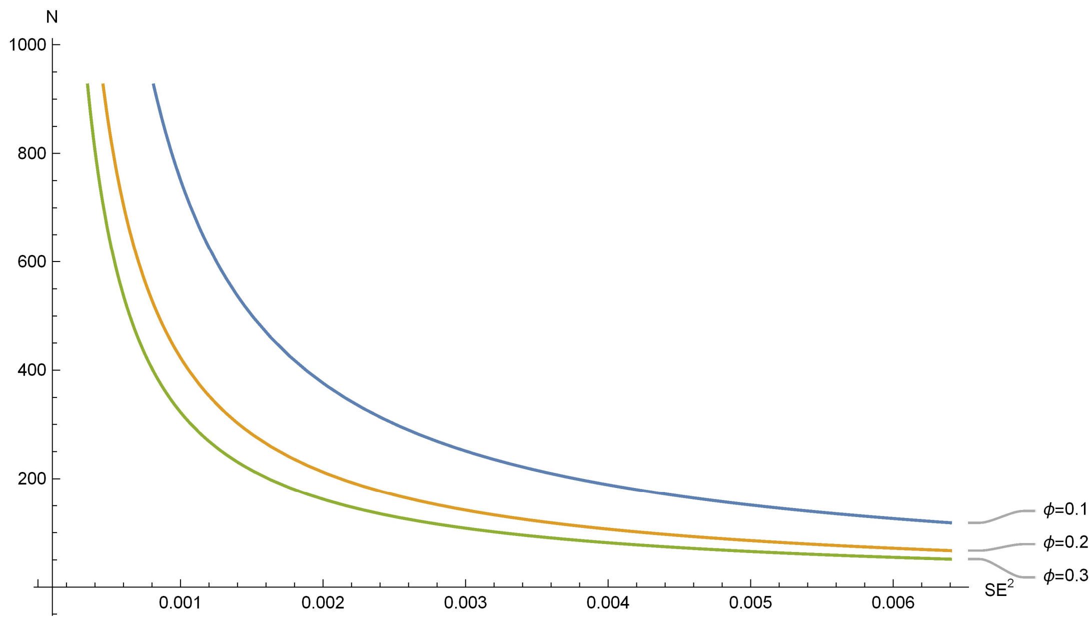
Out[28]=



Estimated sample size *N* as a function of $SE(C)^2$ for different ϕ values.

In[29]:=
```
ϕ1 = sol2 /. {Cs → 0.7, ϕ → 0.1};
ϕ2 = sol2 /. {Cs → 0.7, ϕ → 0.2};
ϕ3 = sol2 /. {Cs → 0.7, ϕ → 0.3};
Plot[{ϕ1, ϕ2, ϕ3}, {SE2, 0.01^2, 0.08^2},
 PlotLabels → {"ϕ=0.1", "ϕ=0.2", "ϕ=0.3"}, AxesLabel → {"SE^2", "N"}, ImageSize → Large]
```

Out[32]=



In[33]:= `sol2`

Out[33]=

$$\left(Cs + Cs^2 - 4\,Cs^3 + 2\,Cs^4 + \sqrt{\left((-1+Cs)\,Cs\,\left((-1+Cs)\,Cs\,(1-2\,(-1+Cs)\,Cs)^2 + 16\,\left(-2+Cs-2\,Cs^3+Cs^4\right)\,SE2\,\phi - 16\,\left(-2+Cs-2\,Cs^3+Cs^4\right)\,SE2\,\phi^2\right)\right)}\right) \Big/ \left(4\,(-2+Cs)\,(1+Cs)\,SE2\,(-1+\phi)\,\phi\right)$$

Further reduction of sol2 by substitution.

sol2 contains repeated subexpressions.

In[34]:=
```
Clear[subexpr0, subexpr1, subexpr2];
subexpr0 = (-1 + Cs);
subexpr1 = 16 (-2 + Cs - 2 Cs^3 + Cs^4) SE2 ϕ;
subexpr2 = -16 (-2 + Cs - 2 Cs^3 + Cs^4) SE2 ϕ^2;
Count[sol2, subexpr0, Infinity]
Count[sol2, subexpr1, Infinity]
Count[sol2, subexpr2, Infinity]
subexpr2 / subexpr1
```

Out[38]= 3

Out[39]= 1

Out[40]= 1

Out[41]= $-\phi$

sol3 is a more compact representation of sol2:

In[42]:=
```
Clear[sol3, α, β];
sol3 = Simplify[
  sol2 /. {(-1 + Cs) → α, -2 + Cs → α - 1, (1 + Cs) → α + 2, subexpr1 → β, subexpr2 → -ϕ β}]
```

Out[43]=
$$\frac{Cs + Cs^2 - 4\,Cs^3 + 2\,Cs^4 + \sqrt{Cs\,\alpha\,\left(Cs\,\alpha\,(1 - 2\,Cs\,\alpha)^2 + \beta - \beta\,\phi\right)}}{4\,SE2\,(-1 + \alpha)\,(2 + \alpha)\,(-1 + \phi)\,\phi}$$

The next replacement for Cs α requires a more explicit replacement rule, as the term occurs within a square root:

In[44]:=
```
Clear[rule1, sol4, subexpr3];
rule1 =
 (Cs α (Cs α (1 - 2 Cs α)^2 + β - β ϕ))^Rational[1, 2] → (μ (μ (1 - 2 μ)^2 + β - β ϕ))^Rational[1, 2]
subexpr3 = Cs (-1 + Cs);
```

Out[45]=
$$\sqrt{Cs\,\alpha\,\left(Cs\,\alpha\,(1 - 2\,Cs\,\alpha)^2 + \beta - \beta\,\phi\right)} \to \sqrt{\mu\,\left(\beta + (1 - 2\,\mu)^2\,\mu - \beta\,\phi\right)}$$

In[47]:=
```
sol4 = sol3 /. rule1
```

Out[47]=
$$\frac{Cs + Cs^2 - 4\,Cs^3 + 2\,Cs^4 + \sqrt{\mu\,\left(\beta + (1 - 2\,\mu)^2\,\mu - \beta\,\phi\right)}}{4\,SE2\,(-1 + \alpha)\,(2 + \alpha)\,(-1 + \phi)\,\phi}$$

Confirm the equivalency of sol2 and sol4:

In[48]:=
```
sol2 /. {Cs → 0.75, ϕ → 0.018, SE2 → 0.02551^2}
sol2 /. {Cs → 0.85, ϕ → 0.01, SE2 → 0.02551^2}
```

Out[48]= 5124.14

Out[49]= 5838.5

In[50]:=
```
sol4 /. {α → subexpr0, β → subexpr1, μ → subexpr3} /.
  {Cs → 0.75, ϕ → 0.018, SE2 → 0.02551^2}
sol4 /. {α → subexpr0, β → subexpr1, μ → subexpr3} /.
  {Cs → 0.85, ϕ → 0.01, SE2 → 0.02551^2}
```

Out[50]= 5124.14

Out[51]= 5838.5

In[52]:=
```
InputForm[sol4]
```

Out[52]//InputForm=
```
(Cs + Cs^2 - 4*Cs^3 + 2*Cs^4 + Sqrt[μ*(β + (1 - 2*μ)^2*μ - β*ϕ)])/(4*SE2*(-1 + α)*(2 + α)*(-1 + ϕ
```