Appendix A: Asymptotic variances of $\widehat{iPCF}$ and $\widehat{iPNF}$ estimated using observed risks in the population

As in previous work for PCF and PNF (Pfeiffer and Gail, 2011), we use an influence function approach for functionals $T(F_n)$ of the empirical distribution function, F_n, to obtain non-model based variance estimates through a first order approximation of T (Huber, 1996),

$$T(F_n) \approx T(F) + \frac{1}{n}\sum_{1}^{n}\psi(y_i). \tag{0.1}$$

The influence functions $\psi(y_i)$ are independent with $E\psi(y) = 0$. Under regularity conditions that assure that the remainder term in the linearization is of the order $o(n^{-1})$, which are satisfied if T is Fréchet or Hadamard differentiable (van der Vaart, 1998),

$$\text{var}\{T(F_n)\} = \frac{1}{n}E\psi^2(y_i). \tag{0.2}$$

The asymptotic normality of $n^{1/2}\{T(F_n) - T(F)\}$ follows from (0.1), (0.2) and the application of the central limit theorem. Under these regularity conditions, the influence functions $\psi(y)$ can be computed as the derivative

$$\psi(y) = \frac{d}{dt}T(F_t)|_{t=0},$$

where $F_t = (1-t)F + t\delta_y$ and δ_y denotes the point mass at the point y.

We previously computed the influence function ψ_L for the Lorenz curve, i.e. $T(F_n) = L_n$ (Pfeiffer and Gail, 2011), given by

$$\psi_L(r,p) = -\frac{1}{\mu}(r-\mu)L(p) - \frac{1}{\mu}\{F^{-1}(p) - r\}I\{r \leqslant F^{-1}(p)\} + \frac{1}{\mu}F^{-1}(p)p - L(p).$$

Thus the influence function of $T(F_t) = -\int_0^{p*} L_t$ and thus $iPCF$ is obtained as

$$\psi_{iPCF}(r) = -\frac{d}{dt}T(F_t)|_{t=0} = -\frac{d}{dt}\Big\{\int_0^{p*} L_t\Big\}|_{t=0} = -\int_0^{p*}\frac{d}{dt}L_t|_{t=0} = \int_0^{p*}\psi_L(p,r)dp. \tag{0.3}$$

$E\psi_{iPCF}F(r) = 0$ as $E\psi_L(y) = 0$.

Similarly, we get the asymptotic variance of $\widehat{iPNF}$ by based on the influence function of PNF, or equivalently, the inverse of the Lorenz curve, L^{-1},

$$\psi_{L^{-1}}(r) = -L^{-1}(p) + \frac{r}{G^{-1}(p)}[p - I\{r \leqslant G^{-1}(p)\}] + I\{r \leqslant G^{-1}(p)\},$$

using a change of variable as

$$\psi_{iPNF}(r) = \int_0^{\gamma_q^*} -\frac{r}{u}\{G(u) - I(y \leqslant u)\} - I(y \leqslant u) + F(u)dG(u). \tag{0.4}$$

Again, $E\psi_{iPNF} = 0$ as $E\psi_{L^{-1}}(r) = 0$.

Appendix B: Asymptotic variances of $\widehat{PCF}$, $\widehat{PNF}$, $\widehat{iPCF}$ and $\widehat{iPNF}$ estimated from case-control data

We utilize a partial influence function approach (Pires and Branco, 2002) that allows to derive the asymptotic properties of functionals of empirical distribution functions for two or more independent populations.

Let $T(K_n, G_m)$ denote a functional of two empirical distribution functions, K_n and G_m, estimated from the independent samples $x_i \sim K, i = 1, \ldots, n$ and $y_j \sim G, j = 1, \ldots, m$ and let $N = m + n$. Then a first order approximation of T yields

$$T(K_n, G_m) \approx T(K, G) + \sum_i \psi_1(x_i)/n + \sum_j \psi_2(y_j)/m. \tag{0.5}$$

The partial influence functions ψ_1 and ψ_2 are independent with $E_K\psi_1(x) = 0$ and $E_G\psi_2(y) = 0$. Again, under regularity conditions the remainder term in the linearization is of the order $o(N^{-1})$ and, letting $\lambda = n/N$, $\sqrt{N}(T(K_n, G_m) - T(K, G)) \to N(0, \text{var}\{T(K_n, G_m)\})$ with

$$N\text{var}\{T(K_n, G_m)\} = \lambda^{-1}E_K\psi_1^2(x) + (1-\lambda)^{-1}E_G\psi_2^2(y). \tag{0.6}$$

Letting $K_t(y) = (1-t)K(y) + t\delta_y$, we compute the influence function ψ_1 for $T(K_t, G)$ as $\psi_1 = \frac{d}{dt}T(K_t, G)|_{t=0}$. Similarly we derive $\psi_2 = \frac{d}{dt}T(K, G_t)|_{t=0}$, with $G_t(y) = (1-t)G + t\delta_y$.

We assume that the distributions G and K are sufficiently smooth so that their densities g and k and thus the density of the population distribution F given by $f = \mu g + (1-\mu)k$ in the points $1-p$ and $1-q$ exist. Letting $\phi_{1-p} = F^{-1}(1-p)$, the partial influence functions for $PCF(p)$ are

$$\psi_1(x) = -(1-\mu)\frac{g(\phi_{1-p})}{f(\phi_{1-p})}\{I(0,\phi_{1-p})(x) - K(\phi_{1-p})\}$$

and

$$\psi_2(y) = -\mu\frac{g(\phi_{1-p})}{f(\phi_{1-p})}\{I(0,\phi_{1-p})(y) - G(\phi_{1-p})\}.$$

The variance terms in (0.6) for PCF are

$$E_K\psi_1^2(x) = (1-\mu)^2\frac{g^2(\phi_{1-p})}{f^2(\phi_{1-p})}K(\phi_{1-p})\{1 - K(\phi_{1-p})\}$$

and

$$E_G\psi_2^2(y) = \mu^2\frac{g^2(\phi_{1-p})}{f^2(\phi_{1-p})}G(\phi_{1-p})\{1 - G(\phi_{1-p})\}.$$

For PNF, letting $\gamma_{1-q} = G^{-1}(1-q)$, the partial influence functions are

$$\psi_1(x) = -(1-\mu)\{I(0,\gamma_{1-q})(x) - K(\gamma_{1-q})\}$$

and

$$\psi_2(y) = (1-\mu)\frac{k(\gamma_{1-q})}{g(\gamma_{1-q})}\{I(0,\gamma_{1-q})(y) - (1-q)\}.$$

The corresponding variance terms are

$$E_K\psi_1^2 = (1-\mu)^2K(\gamma_{1-q})\{1 - K(\gamma_{1-q})\} \text{ and } E_G\psi_2^2 = \left\{(1-\mu)\frac{k(\gamma_{1-q})}{g(\gamma_{1-q})}\right\}^2 q(1-q).$$

For $iPCF(p^*)$ we get, using Leibniz's rule,

$$\psi_1(x) = (1-\mu)G(\phi_{1-p})\{I(0,\phi_{1-p})(x) - K(\phi_{1-p})\} - (1-\mu)\left\{G(x)I(0,\phi_{1-p})(x) - \int_0^{\phi_{1-p}} G(u)dK(u)\right\}$$

and

$$\psi_2(y) = -(1-\mu)K(\phi_{1-p})\{I(0,\phi_{1-p})(y) - G(\phi_{1-p})\} + (1-\mu)\left\{K(y)I(0,\phi_{1-p})(y) - \int_0^{\phi_{1-p}} K(u)dG(u)\right\}$$

with corresponding variance terms

$$E_K\psi_1^2(x) = (1-\mu)^2\Big[G^2(\phi_{1-p})K(\phi_{1-p})\{1-K(\phi_{1-p})\}+\int_0^{\phi_{1-p}} G(x)^2dK(x)-\Big\{\int_0^{\phi_{1-p}} G(x)dK(x)\Big\}^2-$$
$$2G(\phi_{1-p})\{1-K(\phi_{1-p})\}\int_0^{\phi_{1-p}} G(x)dK(x)\Big]$$

$$E_G\psi_2^2(y) = (1-\mu)^2\Big[\int_0^{\phi_{1-p}} K(x)^2dG(x) - \Big\{\int_0^{\phi_{1-p}} K(x)dG(x)\Big\}^2 +$$
$$K^2(\phi_{1-p})G(\phi_{1-p})\{1-G(\phi_{1-p})\} - 2K(\phi_{1-p})\{1-G(\phi_{1-p})\}\int_0^{\phi_{1-p}} K(x)dG(x)\Big]$$

For $iPNF(p^*)$, using Leibniz's rule

$$\psi_1(x) = -(1-\mu)(1-q^*)\Big\{I(0,\gamma_{1-q})(x)-K(\gamma_{1-q})\Big\}+(1-\mu)\Big\{G(x)I(0,\gamma_{1-q})(x)-\int_0^{\gamma_{1-q}} G(u)dK(u)\Big\}$$

and

$$\psi_2(y) = (1-\mu)\Big[K(\gamma_{1-q})\{I(0,\gamma_{1-q})(y)-1+q\} - K(y)I(0,\gamma_{1-q})(y) + \int_0^{\gamma_{1-q}} K(u)dG(u)\Big]$$

and

$$E_K\psi_1^2(x) = (1-\mu)^2\Big[K(\gamma_{1-q})\{1-K(\gamma_{1-q})\}(1-q)^2+\int_0^{\gamma_{1-q}} G^2(u)dK(u)-\Big\{\int_0^{\gamma_{1-q}} G(u)dK(u)\Big\}^2-$$
$$2(1-q)\{1-K(\gamma_{1-q})\}\int_0^{\gamma_{1-q}} G(x)dK(x)\Big]$$

$$E_G\psi_2^2(y) = (1-\mu)^2\Big[K^2(\gamma_{1-q})q(1-q) - 2(1-q)K(\gamma_{1-q})\int_0^{\gamma_{1-q}} K(u)dG(u)+$$
$$\int_0^{\gamma_{1-q}} K^2(u)dG(u) - \Big\{\int_0^{\gamma_{1-q}} K(u)dG(u)\Big\}^2\Big]$$

Appendix C: Asymptotic variances of $\widehat{PCF}$, $\widehat{PNF}$, $\widehat{iPCF}$ and $\widehat{iPNF}$ estimated using risks and outcomes in a cohort

Here we observe $(r_i^F, y_i), i = 1, \ldots, N$ of risks and associated outcomes in a population, that are realizations of the bivariate random variable $(R, Y) \sim F(r, y)$. We now treat $\widehat{PCF}$, $\widehat{PNF}$, $\widehat{iPCF}$

and $\widehat{iPNF}$ as continuous functionals of the bivariate empirical distribution function

$$F_N(r^*, y^*) = \frac{1}{N}\sum_{i=1}^{N} I(r_i \leqslant r^*, y_i \leqslant y^*).$$

To derive the influence functions, we use that the marginal distribution function of the risk R is $F_R(r) = \int_y dF(r,y)$ and the distribution of risk in cases is $G(r) = F(r, y=1)/\mu$, where $\mu = P(y=1)$. As $G_t = 1/\mu_t F_t(r, y=1)$,

$$\frac{d}{dt}G_t(x) = -\frac{1}{\mu}(y-\mu)G(x) + \frac{1}{\mu}\delta(x, y=1) - G(x)$$

and

$$\frac{d}{dt}G_t^{-1}(p) = \frac{1}{\mu g(\gamma_p)}\{-(y-\mu) + I(0,\gamma_p) - p\mu\}.$$

Based on similar linearization arguments as used for the derivations in Appendix A, we obtain that both $\widehat{PCF}$ and $\widehat{PNF}$ based on (r^F, y) are asymptotically normally distributed with variances obtained from the bivariate influence functions

$$\psi_{PCF}(r,y) = \frac{1}{\mu}(y-\mu)G(\phi_{1-p}) + \frac{g(\phi_{1-p})}{f(\phi_{1-p})}\Big\{I(0,\phi_{1-p})(r) - (1-p)\Big\} - \Big\{\frac{1}{\mu}I(0,\phi_{1-p})(r, y=1) - G(\phi_{1-p})\Big\}$$

and

$$\psi_{PNF}(r,y) = -\frac{f_R(\gamma_{1-q})}{g(\gamma_{1-q})}\Big\{-(y-\mu)(1-q) + \frac{1}{\mu}I(0,\gamma_{1-q})(r) - (1-q) - I(0,\gamma_{1-q})(r) + F_R(\gamma_{1-q})\Big\},$$

where $\phi_{1-p} = F_R^{-1}(1-p)$ and $\gamma_{1-q} = G^{-1}(1-q)$.

For $iPCF$, the bivariate influence function is given by

$$\psi_{iPCF}(r,y) = \{I(0,\phi_{1-p})(r) - (1-p)\}G(\phi_{1-p}) + \frac{y-\mu}{\mu}\int_0^{\phi_{1-p}} G(x)dF_R(x) - \int_0^{\phi_{1-p}} \frac{I(x, y=1)(r)}{\mu}dF_R(x) - G(r)I(0,\phi_{1-p})(r) + 2\int_0^{\phi_{1-p}} G(x)dF_R(x),$$

and for $iPNF$, by

$$\psi_{iPNF}(r,y)(q) = -F_R(\gamma_{1-q})\Big\{\frac{y-\mu}{\mu}(1-q) - \frac{1}{\mu}I(0,\gamma_{1-q})(r,y=1) + (1-q)\Big\}$$
$$+ \frac{y-\mu}{\mu}\int_0^{\gamma_{1-q}} F_R(x)dG(x) - \frac{1}{\mu}I(0,\gamma_{1-q})(r,y=1)F_R(r)$$
$$+ 2\int_0^{\gamma_{1-q}} F_R(x)dG(x) - \int_0^{\gamma_{1-q}} I(0,x)(r)dG(x).$$

References

Huber, P.J. (1996). Robust Statistical Procedures. Society for Industrial and Applied Mathematics.

Pfeiffer R.M., Gail M.H. (2011). Two criteria for evaluating risk prediction models, *Biometrics* 67 (3), 1057–1065.

Pires, A. M. and Branco, J. A. (2002). Partial influence functions. ıJournal of Multivariate Analysis 83, 2, 451-468.

van der Vaart, A.W. (1998). Asymptotic Statistics. Cambridge University Press, Cambridge, UK.