data structure. Our principle concern will be with the estimation of causal and non-causal parameters in longitudinal studies in which data are available on all time dependent covariates that predict both (i) subsequent response and (ii) subsequent treatment and/or censoring, so that the data are CAR.

As discussed above, we can use our general estimating function methodology to obtain $n^{1/2}-$ consistent estimators $\widehat{\mu}$ of a smooth parameters μ of a very large non- or semiparametric model for the law F_X of a high-dimensional X from CAR data Y, provided either that the CAR mechanism G is known or that we can correctly specify a lower dimension model for G. Now, in observational studies in which data are missing by happenstance and subjects self-select treatment, even if we are willing to assume the data are CAR, nonetheless the density g of Y given X will not be known; further, we cannot be certain the lower dimensional model we assume for g is correct. Thus we cannot be assured that our estimator $\widehat{\mu}$ is consistent. Because of this uncertainty, we might choose to specify a lower dimensional (say a fully parametric) working submodel $f(X;\mu,\eta)$ of our large non- or semiparametric model for F_X and then estimate the finite dimensional parameters (μ,η) based on the data Y by parametric maximum likelihood. The difficulty with this approach is that, if the parametric submodel $f(X;\mu,\eta)$ is misspecified, the parametric MLE of μ will be inconsistent. However, because of the curse of dimensionality, we cannot obtain estimators with reasonable finite sample performance if we do not place additional modelling restrictions on either the CAR mechanism G or on our large non- or semiparametric model for F_X. Hence, the best that can be hoped for is to find a doubly robust estimator. An estimator is doubly robust (equivalently, doubly protected) if it is consistent asymptotically normal (CAN) under the assumption of CAR when either (but not necessarily both) a lower dimensional model for G or a lower dimensional model for F_X is correct. A doubly robust estimator is locally semiparametric efficient (LSE) if it is the asymptotically most efficient doubly robust estimator of μ when both the lower dimensional models for G and F_X happen to be correct.

It turns out that, as discussed by Scharfstein, Rotnitzky, and Robins (1999), Neugebauer, van der Laan (2002), and later in this chapter, with a little care, we can guarantee that the aforementioned LSE estimator of μ in the semiparametric model that assumes a correct lower dimensional model for g is actually a LSE doubly robust estimator. Specifically the aforementioned LSE estimator of μ based on our general estimating function methodology actually depends not only on an estimate of g but also on an estimate of the law F_X. Further, because of the curse of dimensionality, it is necessary that F_X be estimated using a lower dimensional working submodel. If the model for g is correct, our LSE estimator is CAN for μ regardless of whether our working submodel for F_X is correct. However, if this submodel is correct, our estimator of μ attains the efficiency bound for

the semiparametric model that assumes a correct lower dimensional model for G.

Now, if we take care that our estimate of F_X under the lower dimensional submodel is the MLE and thus depends only on the F_X part of the likelihood, then the estimator of μ considered in the previous paragraph is actually a LSE doubly robust estimator; in particular, it is CAN for μ even if the model for g is incorrect, provided the lower dimensional (say, parametric) submodel for F_X is correct. Even more surprisingly, Scharfstein, Rotnitzky, and Robins (1999) and Neugebauer, and van der Laan (2002) show that, when the lower dimensional (say, parametric) model for F_X happens to be correct, this doubly robust estimator of μ, like the parametric MLE, may be CAN, even when μ is not identified under the observed data model in which the true density g of Y given X is completely known. Non-identifiability of μ in this latter model occurs when the support set for Y at each value of X under the known density g is very small. See Section 1.6 for details. Thus, in CAR models, it is best to (i) simultaneously model the coarsening (i.e., censoring and/or treatment) mechanism and the law of the full data with lower dimensional models, (ii) estimate them both by maximum likelihood separately from the two parts of the likelihood, and (iii) finally obtain a LSE doubly robust estimator with our general estimating function methodology (see e.g., Scharfstein, Rotnitzky, and Robins, 1999, and Robins, 2000). Yu, and van der Laan (2002) implement this strategy and provide explicit algorithmic and computational suggestions: See Section 1.6. This strategy, however, is not always computationally feasible; in that case, alternative approaches to doubly robust estimation are available as developed in Robins (2000) and Tchetgen and Robins (2002). See Sections 3.5 and 6.4.

Certain semiparametric models in addition to CAR censored data models also admit doubly robust estimators. As far as we are aware, Brillinger (1983) was the first to call attention to and provide examples of DR-like estimators. Other examples are given by Ruud (1983, 1986), Duan and Li (1987, 1991), Newey (1990), Robins, Mark and Newey (1992), Ritov and Robins (1997), and Lipsitz and Ibrahim (1999). Scharfstein, Rotnitzky, Robins (1999), Robins (2000), and Neugebauer, van der Laan (2002) went beyond individual examples to provide a broad theory of double robustness in missing data and counterfactual causal inference models in which the data was CAR. Robins, Rotnitzky, and Van der Laan (2000) extended the results in Scharfstein, Rotnitzky, Robins (1999), and Robins (2000) to cover DR estimation in any model in which locally variation-independent (possibly infinite-dimensional) parameters κ and γ index the law of the observed data, the likelihood factorized as $L(\kappa,\gamma) = L_1(\kappa)\,L_2(\gamma)$, and the smooth finite-dimensional parameter $\mu(\kappa,\gamma) = \mu(\kappa)$ only depended on κ. All of the above mentioned examples are special cases of the general results in Robins, Rotnitzky, van der Laan (2000). Robins and Rotnitzky (2001), in the most comprehensive investigation of double robustness to date, provide