Chapter 25
Probability of Success of an In Vitro Fertilization Program

Antoine Chambaz

About 9 to 15% of couples have difficulty in conceiving a child, i.e., do not conceive within 12 months of attempting pregnancy (Boivin et al. 2007). In response to subfertility, assisted reproductive technology has developed over the last 30 years, resulting in in vitro fertilization (IVF) techniques (the first "test-tube baby" was born in 1978). Nowadays, more than 40,000 IVF cycles are performed each year in France and more than 63,000 in the USA (Adamson et al. 2006). Yet, how to quantify the success in assisted reproduction still remains a matter of debate. One could, for instance, rely on the number of pregnancies or deliveries per IVF cycle. However, an IVF program often consists of several successive IVF cycles. So, instead of considering each IVF cycle separately, one could rather rely on an evaluation of the whole program. IVF programs are emotionally and physically burdensome. Providing the patients with the most adequate and accurate measure of success is therefore an important issue that we propose to address in this chapter.

25.1 The DAIFI Study

Our contribution is based on the French Devenir Après Interruption de la FIV (DAIFI) study (Soullier et al. 2008; de la Rochebrochard et al. 2008, 2009). In France, the four first IVF cycles are fully reimbursed by the national health insurance system. Therefore, as in the previous references, we conclude that the most adequate measure of success for French couples is the probability of delivery (resulting from embryo transfer) during the course of a program of at most four IVF cycles. We will refer to this quantity as the probability of success of a program of at most four IVF cycles, or even sometimes as the probability of success.

Data were provided by two French IVF units (Cochin in Paris and Clermont-Ferrand, a medium-sized city in central France). All women who followed their first IVF cycle in these units between 1998 and 2002 and who were under 42 at the start of the cycle were included. Women over 42 were not included, unless they had a

M.J. van der Laan and S. Rose, *Targeted Learning: Causal Inference for Observational and Experimental Data*, Springer Series in Statistics, DOI 10.1007/978-1-4419-9782-1_25,

normal ovarian reserve and a specific IVF indication. For every enrolled woman, the data were mainly the attended IVF unit and the woman's date of birth, and for each IVF cycle, its start date, number of oocytes harvested, number of embryos transferred or frozen, indicators of pregnancy, and successful delivery (for a comprehensive description, see de la Rochebrochard et al. 2009). Data collection was discontinued after the woman's fourth IVF cycle. Since the first four IVF cycles are fully reimbursed, it is reasonable to assume that economic factors do not play a role in the phenomenon of interest. Specifically, whether a couple will abandon the IVF program mid-course without a successful delivery or undergo the whole program does not depend on economic factors (on the contrary, if the IVF cycles were not fully reimbursed, then disadvantaged couples would likely abandon the program mid-course more easily). Furthermore, successive IVF cycles occur close together in time: hence, the sole age at the start of the first IVF cycle is a relevant summary of the successive ages at the start of each IVF cycle during the program. Likewise, we make the assumption that the number of embryos transferred or frozen during the first IVF cycle is a relevant summary of the successive number of harvested oocytes and transferred or frozen embryos associated to each IVF cycle (i.e., a relevant summary measure of the couple's fertility during the program). Relaxing this assumption will be considered in future work.

Estimating the probability of success of a program of at most four IVF cycles is not easy due to couples who abandon the IVF program mid-course without a successful delivery. Moreover, since those couples have a smaller probability of having a child than couples that undergo the whole program, it would be wrong to ignore the right censoring, and simply count the proportion of successes (Soullier et al. 2008), even if the decision to abandon the program is not informed by any relevant factors. In addition, it seems likely that some of the baseline factors, such as baseline fertility, might be predictive of the dropout time (measured on the discrete scale of number of IVF cycles): in statistical terms, we expect that the right-censoring mechanism will be informative.

Three approaches to estimating the probability of success of a program of at most four IVF cycles are considered in Soullier et al. (2008). The most naive approach estimates the probability of success as the ratio of the number of deliveries successive to the first IVF cycle to the total number of enrolled women, yielding a point estimate of 37% and a 95% confidence interval given by $(0.35, 0.38)$. This first approach obviously overlooks a lot of information in the data.

A second approach is a standard nonparametric survival analysis based on the Kaplan–Meier estimator. Specifically, Soullier et al. (2008) compute the Kaplan–Meier estimate S_n of the survival function $t \mapsto P(T \geq t)$, where T denotes the number of IVF cycles attempted after the first one till the first successful delivery. The observed data structure is represented as $(\min(T, C), I(T \leq C))$, with C the right-censoring time. The estimated probability of success is given by $1 - S_n(3)$. This method resulted in an estimated probability of success equal to 52% and a 95% confidence interval $(0.49, 0.55)$. This much more sensible approach still neglects the baseline covariates and thus assumes that a woman's decision to abandon the program is not informed by relevant factors that predict future success, such as those

measured at baseline. One could argue that this method should provide an estimated upper bound on the probability of success.

Actually, formulating the problem of interest in terms of survival analysis is a sensible option, and indeed the methods in Part V for right-censored data can be employed to estimate the survival function of T. In Sect. 25.6, we discuss the equivalence between our approach presented here and the survival analysis approach.

In order to improve the estimate of the probability of success, Soullier et al. (2008) finally resort to the so-called multiple imputation methodology (Schafer 1997; Little and Rubin 2002). Based on iteratively estimating missing data using the past, this third approach leads to a point estimate equal to 46%, with a 95% confidence interval given by (0.44, 0.48).

The three methods that we summarized either answer only partially the question of interest (naive approach and nonparametric Kaplan–Meier analysis) or suffer from bias due to reliance on parametric models (multiple-imputation approach). We expose in the next sections how the TMLE methodology paves the way to solving this delicate problem with great consideration for theoretical validity.

25.2 Data, Model, and Parameter

The observed data structure is longitudinal:

$$O = (L_0, A_0, L_1, A_1, L_2, A_2, L_3 = Y),$$

where $L_0 = (L_{0,1}, L_{0,2}, L_{0,3}, L_{0,4})$ denote the baseline covariates and $L_{0,1}$ indicates the IVF center, $L_{0,2}$ indicates the age of the woman at the start of the first IVF cycle, $L_{0,3}$ indicates the number of embryos transferred or frozen at the first IVF cycle, and $L_{0,4}$ indicates whether the first IVF cycle is successful, i.e., yields a delivery, ($L_{0,4} = 1$) or not ($L_{0,4} = 0$). For each $1 \leq j \leq 3$, A_{j-1} indicates whether the woman completes her jth IVF cycle ($A_{j-1} = 1$) or not ($A_{j-1} = 0$) this also encodes for dropout, and L_j indicates whether the jth IVF cycle is successful ($L_j = 1$) or not ($L_j = 0$). The longitudinal data structure becomes degenerate after a time point t at which either the woman abandons the program ($A_t = 0$) or has a successful IVF cycle ($L_t = 1$ for some t). By encoding convention, the data structure O is constrained as follows. (1) If $A_{j-1} = 0$ for some $1 \leq j < 3$, then $A_j = \ldots = A_2 = 0$ and $L_j = \ldots = L_3 = Y = 0$. (2) If $L_{0,4} = 1$, then $L_1 = \ldots = L_3 = Y = 1$ and $A_0 = \ldots = A_2 = 1$, and similarly if $L_j = 1$ for some $1 \leq j < 3$, then $L_{j+1} = \ldots = L_3 = Y = 1$ and $A_j = \ldots = A_2 = 1$, too.

The true data generating distribution of O is denoted by P_0. We assume that the following positivity assumption holds: P_0–almost surely, for each $0 \leq j \leq 2$:

$$0 < P_0\left(A_j = 1 \mid L_{1:j} = 0_{1:j}, A_{0:j-1} = 1_{0:j-1}, L_{0,4} = 0, (L_{0,1}, L_{0,2}, L_{0,3})\right),$$

with notation $x_{i:j} = (x_i, \ldots, x_j) \in \mathbb{R}^{j-i+1}$ for $i \leq j$, and the obvious convention $x_{1:0} = A_{1:0} = L_{1:0} = \emptyset$. This assumption states that, for each $0 \leq j \leq 2$, conditionally

on observing a woman who already went through $(j+1)$ unsuccessful IVF cycles, it cannot be certain, based on past information $(L_{0:j}, A_{0:j-1})$, that a $(j+2)$-th IVF cycle will not be attempted. As discussed in Chap. 10, positivity can be tested from the data.

We see P_0 as an element of the statistical model $\mathcal{M}$ of all possible probability distributions of O (satisfying, in particular, the constraints imposed by the encoding convention and positivity assumption). Set $\mathcal{M}$ is large because a model should reflect only true knowledge and because we lack any meaningful knowledge about the true data-generating distribution P_0 that would allow us to enforce further restrictions.

The parameter mapping of interest, $\Psi(P)$, is the following explicit functional of a candidate data-generating distribution P:

$$\Psi(P) = E_P\Bigg(\sum_{\ell_{1:2} \in \{0,1\}^2} P(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = \ell_{1:2}, L_0) \\ \times P(L_2 = \ell_2 \mid A_{0:1} = 1_{0:1}, L_1 = \ell_1, L_0) \times P(L_1 = \ell_1 \mid A_0 = 1, L_0) \Bigg), \quad (25.1)$$

where, by convention, for each $2 \leq j \leq 3$,

$$P\Big(L_j = \ell_j \mid A_{0:j-1} = 1_{0:j-1}, L_{1:j-1} = \ell_{1:j-1}, L_{0,4} = \ell_{0,4}, (L_{0,1}, L_{0,2}, L_{0,3})\Big) = 0, \quad (25.2)$$

whenever the event $[(L_{0,4}, A_0, L_1, A_1, L_2, A_2, L_3) = (\ell_{0,4}, 1, \ell_1, 1, \ell_2, 1, \ell_3)]$ does not meet the encoding constraints. The objective of this chapter is to provide a point estimate and a 95% confidence interval for the statistical target parameter $\psi_0 = \Psi(P_0)$.

The parameter of interest $\psi_0 = \Psi(P_0)$ in (25.1) is a well-defined statistical parameter on the nonparametric statistical model $\mathcal{M}$. Its causal interpretation is enlightening, and can be derived in two different frameworks. Furthermore, it is not necessary to rely on the causal interpretation to justify the scientific interest of ψ_0 as a pure statistical estimand.

The first causal interpretation of (25.1) relies on the so-called counterfactual framework. Let us first introduce the set $\mathcal{A}$ of all possible realizations of $A_{0:2}$: $\mathcal{A} = \{(0,0,0), (1,0,0), (1,1,0), (1,1,1)\}$. By Theorem 2.1 in Yu and van der Laan (2002), there exists an *explicit* construction $L_a = f_a(O, P_0)$, $a \in \mathcal{A}$, involving augmenting the probability space with an independent draw of uniformly distributed random variables and quantile-quantile functions for discrete random variables, so that we have consistency, P_0–almost surely, $O = (A, L_A)$, and randomization, for all $0 \leq j \leq 2$, A_j, is independent of $X = (L_a : a \in xA)$ conditionally on the observed data history $(L_{0:j}, A_{0:j-1})$. Let $Y_{(1,1,1)}$ denote the last coordinate of $L_{(1,1,1)}$: $Y_{(1,1,1)} = 1$ if and only if the woman finally gives birth after four IVF cycles (the IVF program being interrupted if the woman gives birth mid-course).

Theorem 3.1 of Yu and van der Laan (2002), also guarantees that

$$\begin{aligned}\Pr(Y_{(1,1,1)} = 1) = E_{P_0}\Bigg(\sum_{\ell_{1:2} \in \{0,1\}^2} & P_0 \left(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = \ell_{1:2}, L_0\right) \\ & \times P_0 \left(L_2 = \ell_2 \mid A_{0:1} = 1_{0:1}, L_1 = \ell_1, L_0\right) \\ & \times P_0 \left(L_1 = \ell_1 \mid A_0 = 1, L_0\right) \Bigg). \end{aligned} \tag{25.3}$$

This equality is an example of the g-computation formula. It relates the probability distribution of $Y_{(1,1,1)}$ to the probability distribution of the observed data structure O. In addition, it teaches us that $\psi_0 = \Psi(P_0)$ can be interpreted (at the cost of a weak assumption) as the probability of a successful outcome after four IVF cycles (the IVF program being interrupted if the woman gives birth mid-course). Note that when $L_{0,4} = 1$, the sum has only one nonzero term, whereas it has three nonzero terms when $L_{0,4} = 0$.

We need to emphasize that the counterfactuals whose existence is guaranteed by these theorems in Yu and van der Laan (2002), and Gill and Robins (2001) are not necessarily interesting nor have an interpretation that is causal *in the real world*. The structural equations framework that we present hereafter makes the definition of counterfactuals explicit and truly causal since they correspond with intervening on the system of equations (Chap. 2). Alternatively, as in the Neyman–Rubin counterfactual framework discussed in Chap. 21, one defines the counterfactuals in terms of an experiment, and one assumes the consistency and randomization assumption with respect to these user-supplied definitions of the counterfactuals.

It is possible, at the cost of untestable (and stronger) assumptions, to provide another interpretation of (25.1). This second interpretation is at the core of Pearl (2009). It is of course compatible with the previous one. Let us assume that the random phenomenon of interest has *no unmeasured confounders*. A causal graph is equivalent to the following system of structural equations: there exist ten independent random variables $(U^1_{L_0}, \ldots, U^4_{L_0}, U_{A_0}, U_{L_1}, \ldots, U_{A_2}, U_{L_3})$ and ten deterministic functions $(f^1_{L_0}, \ldots, f^4_{L_0}, f_{A_0}, f_{L_1}, \ldots, f_{A_2}, f_{L_3})$ such that

$$\begin{cases} L_{0,1} = f^1_{L_0}(U^1_{L_0}), \\ L_{0,2} = f^2_{L_0}(L_{0,1}, U^2_{L_0}), \\ L_{0,3} = f^3_{L_0}(L_{0,2}, L_{0,1}, U^3_{L_0}), \\ L_{0,4} = f^4_{L_0}(L_{0,3}, L_{0,2}, L_{0,1}, U^4_{L_0}), \\ \quad \text{and for every } 0 \leq j \leq 2, \\ A_j = f_{A_j}(L_{0:j-1}, A_{0:j-1}, U_{A_j}), \\ L_{j+1} = f_{L_{j+1}}(A_{0:j}, L_{0:j}, U_{L_{j+1}}). \end{cases} \tag{25.4}$$

One can intervene upon this system by setting the intervention nodes $A_{0:2}$ equal to some values $a_{0:2} \in \mathcal{A}$. Formally, this simply amounts to substituting the equality $A_j = a_j$ to $A_j = f_{A_j}(L_{0:j}, A_{0:j-1}, U_{A_j})$ for all $0 \leq j \leq 2$ in (25.4). This yields a new causal graph, the so-called graph under intervention $A_{0:2} = a_{0:2}$. The intervened, new, causal graph or system of structural equations describes how $Y = L_3$ is randomly generated under this intervention. Under the intervention $A_{0:2} = a_{0:2}$, this

last (chronologically speaking) random variable is denoted by $Y_{a_{0:2}}$, naturally using the same notation. Moreover, it is known (see, for instance, Robins 1986, 1987a; Pearl 2009) that the g-computation formula (25.3) also holds in this nonparametric structural equation model framework, relating the probability distribution of $Y_{(1,1,1)}$ to the probability distribution of the observed data structure O.

Finally, even if one is not willing to rely on the causal assumptions in the SCM, and one is also not satisfied with the definition of an effect in terms of explicitly constructed counterfactuals, there is still a way forward. Assuming that the time ordering of observed variables $L_{0:4}$ and $A_{0:2}$ is correct (which it indeed is), the target parameter still represents an effect of interest aiming to get as close as possible to a causal effect as the data allow. In any case, $\psi_0 = \Psi(P_0)$ is a well-defined effect of an intervention on the distribution of the data, that can be interpreted as a variable importance measure (Chaps. 4, 22, and 23).

25.3 The TMLE

It can be shown that Ψ is a pathwise differentiable parameter (Appendix A). Therefore the theory of semiparametric models applies, providing a notion of asymptotically efficient estimation and, in particular, its key ingredient, the efficient influence curve. The TMLE procedure takes advantage of the pathwise differentiability and related properties in order to build an asymptotically efficient substitution estimator of $\psi_0 = \Psi(P_0)$.

Let $L_0^2(P)$ denote the set of measurable functions s mapping the set $\mathcal{O}$ (where the observed data structure takes its values) to $\mathbb{R}$, such that $Ps = 0$ and $Ps^2 < \infty$ [we recall that $P\varphi$ is shorthand notation for $E_P\varphi(O)$ for any $\varphi \in L^1(P)$]. A fluctuation model $\{P(\epsilon) : |\epsilon| < \eta\} \subset \mathcal{M}$ is a one-dimensional parametric model such that $P(0) = P$. Its score at $\epsilon = 0$ is $s \in L_0^2(P)$ if the derivative at $\epsilon = 0$ of the log-likelihood $\epsilon \mapsto \log P(\epsilon)(O)$ equals $s(O)$:

$$\frac{\partial}{\partial \epsilon} \log P(\epsilon)(O)\Big|_{\epsilon=0} = s(O).$$

As presented in Chap. 25 of van der Vaart (1998), the functional Ψ is pathwise differentiable at $\mathcal{M}$ with respect to $L_0^2(P)$ if there exists $D \in L_0^2(P)$ such that, for any fluctuation model $\{P(\epsilon) : |\epsilon| < \eta\}$ with score s, the function $\epsilon \mapsto \Psi(P(\epsilon))$ is differentiable at $\epsilon = 0$, with

$$\frac{\partial}{\partial \epsilon} \Psi(P(\epsilon))\Big|_{\epsilon=0} = PsD.$$

In the context of this chapter, if such a $D \in L_0^2(P)$ exists, then it is called the efficient influence curve. Remarkably, the asymptotic variance of a regular estimator of the pathwise differentiable parameter $\Psi(P)$ is lower-bounded by the variance of the efficient influence curve. For that reason in particular, it is important to determine if Ψ is pathwise differentiable and, if it is, to derive its efficient influence curve.

Let us introduce the shorthand notation $Q(L_0; P) = P(L_0)$, $Q(L_1 \mid A_0, L_0; P) = P(L_1 \mid A_0, L_0)$, $Q(L_2 \mid A_{0:1}, L_{0:1}; P) = P(L_2 \mid A_{0:1}, L_{0:1})$, $Q(Y \mid A_{0:2}, L_{0:2}; P) = P(Y \mid A_{0:2}, L_{0:2})$, $g(A_0 \mid X; P) = P(A_0 \mid L_0)$, $g(A_{0:1} \mid X; P) = g(A_0 \mid X; P) \times P(A_1 \mid L_{0:1}, A_0)$, and $g(A_{0:2} \mid X; P) = g(A_{0:1} \mid X; P) \times P(A_2 \mid L_{0:2}, A_{0:1})$. The likelihood $P(O)$ can be represented as

$$P(O) = \prod_{j=0}^{3} Q(L_j \mid A_{0:j-1}, L_{0:j-1}; P)$$
$$\times \prod_{j=0}^{2} g(A_j = 1 \mid A_{0:j-1}, L_{0:j}; P)^{A_j} (1 - g(A_j = 1 \mid A_{0:j-1}, L_{0:j}; P))^{1-A_j}$$

and thus factorizes as $P = Qg$. The parameter of interest at $P = Qg$ can be straightforwardly expressed as a function of Q:

$$\Psi(P) = E_P\Bigg(\sum_{\ell_{1:2} \in \{0,1\}^2} Q(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = \ell_{1:2}, L_0; P)$$
$$\times Q(L_2 = \ell_2 \mid A_{0:1} = 1_{0:1}, L_1 = \ell_1, L_0; P) \times Q(L_1 = \ell_1 \mid A_0 = 1, L_0; P) \Bigg).$$

Note that the outer expectation is with respect to the probability distribution $Q(L_0; P)$ of the baseline covariates L_0.

The following proposition states that Ψ is pathwise differentiable and it presents its efficient influence curve at $P \in \mathcal{M}$.

Proposition 25.1. *The functional Ψ is pathwise differentiable at every $P \in \mathcal{M}$. The efficient influence curve $D^*(\cdot \mid P)$ at $P \in \mathcal{M}$ is written*

$$D^*(\cdot \mid P) = \sum_{j=0}^{3} D_j^*(\cdot \mid P),$$

where

$$\begin{aligned} D_0^*(O \mid P) &= E_P(Y_{(1,1,1)} \mid L_0) - \Psi(P) \\ &= P(L_1 = 1 \mid A_0 = 1, L_0) \\ &\quad + P(L_1 = 0 \mid A_0 = 1, L_0) \times P(L_2 = 1 \mid A_{0:1} = 1_{0:1}, L_1 = 0, L_0) \\ &\quad + P(L_1 = 0 \mid A_0 = 1, L_0) \times P(L_2 = 0 \mid A_{0:1} = 1_{0:1}, L_1 = 0, L_0) \\ &\quad \times P(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = 0_{1:2}, L_0) - \Psi(P), \end{aligned}$$

$$\begin{aligned} D_1^*(O \mid P) &= \frac{I(A_0 = 1)}{g(A_0 = 1 \mid X; P)} \times (L_1 - P(L_1 = 1 \mid A_0, L_0)) \\ &\quad \times \{E_{Q_0}(Y_{(1,1,1)} \mid L_0, A_0 = 1, L_1 = 1) - E_{Q_0}(Y_{(1,1,1)} \mid L_0, A_0 = 1, L_1 = 0)\} \\ &= \frac{I(A_0 = 1)}{g(A_0 = 1 \mid X; P)} \times (1 - P(L_2 = 1 \mid A_{0:1} = 1_{0:1}, L_1 = 0, L_0) \end{aligned}$$

$$
\begin{aligned}
&\times P(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = (0,1), L_0) \\
&- P(L_2 = 0 \mid A_{0:1} = 1_{0:1}, L_1 = 0, L_0) \\
&\times P(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = 0_{1:2}, L_0)) \\
&\times (L_1 - P(L_1 = 1 \mid A_0, L_0)) ,
\end{aligned}
$$

$$
\begin{aligned}
D_2^*(O \mid P) &= \frac{I(A_{0:1} = 1_{0:1})}{g(A_{0:1} = 1_{0:1} \mid X; P)} \times (L_2 - P(L_2 = 1 \mid A_{0:1}, L_{0:1})) \\
&\times \{E_{Q_0}(Y_{(1,1,1)} \mid L_{0:1}, A_{0:1} = 1, L_2 = 1) - E_{Q_0}(Y_{(1,1,1)} \mid L_{0:1}, A_{0:1} = 1, L_2 = 0)\} \\
&= \frac{I(A_{0:1} = 1_{0:1})}{g(A_{0:1} = 1_{0:1} \mid X; P)} \times (P(Y = 1 \mid A_{0:2} = 1_{0:2}, L_2 = 1, L_{0:1}) \\
&\quad - P(Y = 1 \mid A_{0:2} = 1_{0:2}, L_2 = 0, L_{0:1})) \\
&\quad \times (L_2 - P(L_2 = 1 \mid A_{0:1}, L_{0:1})) ,
\end{aligned}
$$

$$
D_3^*(O \mid P) = \frac{I(A_{0:2} = 1_{0:2})}{g(A_{0:2} = 1_{0:2} \mid X; P)} \times (Y - P(Y = 1 \mid A_{0:2}, L_{0:2})) ,
$$

and the latter equalities involve convention (25.2). Furthermore, the efficient influence curve $D^*(\cdot \mid P)$ is double robust: if $P_0 = Q_0 g_0$ and $P = Qg$, then

$$
E_{P_0} D^*(O \mid P) = 0 \quad \textit{implies} \quad \Psi(P) = \Psi(P_0)
$$

if either $Q = Q_0$ or $g = g_0$.

The theory of semiparametric models teaches us that the asymptotic variance of any regular estimator of ψ_0 is lower-bounded by the variance of the efficient influence curve, $E_{P_0} D^*(O \mid P_0)^2$. A regular estimator of ψ_0 having as limit distribution the mean-zero Gaussian distribution with variance $E_{P_0} D^*(O \mid P_0)^2$ is therefore said to be asymptotically efficient.

25.3.1 TMLE Procedure

We assume that we observe n independent copies $O^{(1)}, \ldots, O^{(n)}$ of the observed data structure O. The TMLE procedure takes advantage of the pathwise differentiability of the parameter of interest and bends an initial estimator, obtained as a substitution estimator $\Psi(P_n^0)$, into an updated substitution estimator $\Psi(P_n^*)$ (with P_n^* an update of P_n^0), which enjoys better properties.

Initial estimate. We start by constructing an initial estimate P_n^0 of the distribution P_0 of O, which could also be used to construct an initial estimate $\psi_n^0 = \Psi(P_n^0)$. The initial estimator of the probability distribution of the baseline covariates will be defined as the empirical probability distribution of $L_0^{(i)}$, $i = 1, \ldots, n$. The initial estimate of the other factors of P_n^0 is obtained by super learning, using the log-likelihood loss

function for each of the binary conditional distributions in Q_0.

Updating the initial estimate. The optimal theoretical properties enjoyed by a super learner P_n^0 as an estimator of P_0 do not necessarily translate into optimal properties of $\Psi(P_n^0)$ as an estimator of the parameter of interest $\psi_0 = \Psi(P_0)$. In particular, writing $P_n^0 = Q_n^0 g_n^0$, due to the curse of dimensionality, $\Psi(Q_n^0)$ may still be overly biased due to an optimized tradeoff in bias and variance with respect to the infinite-dimensional parameter Q_0 instead of $\Psi(Q_0)$ itself.

The second step of the TMLE procedure stretches the initial estimate P_n^0 in the direction of the targeted parameter of interest, through a maximum likelihood step. If the initial estimate $\Psi(P_n^0)$ is biased, then this step removes all asymptotic bias for the target parameter whenever the g-factor of P_0, g_0, is estimated consistently: in fact, it maps an inconsistent $\Psi(P_n^0)$ into a consistent TMLE of ψ_0. Hence, the resulting updated estimator is said to be double robust: it is consistent if the initial first-stage estimator of the Q-factor of P_0, Q_0, is consistent or if the g-factor of P_0, g_0, is consistently estimated.

Let's now describe the specific TMLE. We first fluctuate P_n^0 with respect to the conditional distribution of Y given its past $(A_{0:2}, L_{0:2})$, i.e., construct a fluctuation model $\{P_n^0(\epsilon) : |\epsilon| < \eta\}$ through P_n^0 at $\epsilon = 0$ whose score at $\epsilon = 0$ is $D_3^*(\cdot \mid P_n^0)$. Fit ϵ with maximum likelihood. This yields an intermediate update $P_n^0(\epsilon_n^0)$ of P_n^0, which we denote by P_n^1. Then, iteratively from $j = 2$ to $j = 1$, we fluctuate P_n^{3-j} with respect to the conditional distribution of L_j given its past $(A_{0:j-1}, L_{0:j-1})$, using a fluctuation model $\{P_n^{3-j}(\epsilon) : |\epsilon| < \eta\}$ through P_n^{3-j} at $\epsilon = 0$ whose score at $\epsilon = 0$ is $D_j^*(\cdot \mid P_n^{3-j})$, and fitting ϵ with maximum likelihood. This produces a final estimate P_n^* of P_0 that is targeted toward the parameter of interest. The TMLE of ψ_0 is the corresponding substitution estimator $\psi_n^* = \Psi(P_n^*)$.

This TMLE corresponds with selecting the log-likelihood loss function $L(P) = -\log P$ and selecting a parametric model $\{P(\epsilon) : \epsilon\}$, ϵ multivariate, a separate ϵ-component for each factor of $Q(\cdot \mid P)$, no fluctuation of $g(\cdot \mid P)$, and using the recursive backwards MLE-updating algorithm that starts at the last factor and ends at first factor (as originally presented in van der Laan 2010a).

In this simple setting, the construction of the aforementioned fluctuations is easy. It is, for instance, possible to select as parametric fluctuation working models simple univariate logistic regression models. Indeed, let us introduce the so-called *clever covariate* for fluctuating the conditional distribution of Y under P_n^0, the last factor of $Q_n^0 = Q(\cdot \mid P_n^0)$, as $H_{n,3}^* = I(A_{0:2} = 1_{0:2})/g(A_{0:2} = 1_{0:2} \mid X; P_n^0)$. It is straightforward to check that the fluctuation model

$$P_n^0(\epsilon)(O) = \operatorname{expit}\left(\operatorname{logit} Q(Y \mid A_{0:2}, L_{0:2}; P_n^0) + \epsilon H_{n,3}^*\right)^Y \times \left[1 - \operatorname{expit}\left(\operatorname{logit} Q(Y \mid A_{0:2}, L_{0:2}; P_n^0) + \epsilon H_{n,3}^*\right)\right]^{(1-Y)} \times P_n^0(A_{0:2}, L_{0:2})$$

goes through P_n^0 at $\epsilon = 0$, with a score at $\epsilon = 0$ equal to $D_3^*(\cdot; P_n^0)$. Let $\epsilon_{n,3}$ denote the maximum likelihood estimate in the model, and let $P_n^1 = P_n^0(\epsilon_{n,3})$ be the update of P_n^0. Likewise, let us introduce the second clever covariate for fluctuating the

conditional distribution of L_2 under P_n^1, the "next" factor of $Q_n^1 = Q(\cdot \mid P_n^1)$:

$$H_{n,2}^* = \frac{I(A_{0:1} = 1_{0:1})}{g(A_{0:1} = 1_{0:1} \mid X; P_n^1)} \times \Big(P_n^1(Y = 1 \mid A_{0:2} = 1_{0:2}, L_2 = 1, L_{0:1}) - P_n^1(Y = 1 \mid A_{0:2} = 1_{0:2}, L_2 = 0, L_{0:1})\Big)$$

and the related fluctuation model

$$\begin{aligned} P_n^1(\epsilon)(O) = {} & Q(Y \mid A_{0:2}, L_{0:2}; P_n^1) \times g(A_2 = 1 \mid X; P_n^1)^{A_2}(1 - g(A_2 = 1 \mid X; P_n^1))^{1-A_2} \\ & \times \operatorname{expit}\Big(\operatorname{logit} Q(L_2 \mid A_{0:1}, L_{0:1}; P_n^1) + \epsilon H_{n,2}^*\Big)^{L_2} \\ & \times \Big[1 - \operatorname{expit}\Big(\operatorname{logit} Q(L_2 \mid A_{0:1}, L_{0:1}; P_n^1) + \epsilon H_{n,2}^*\Big)\Big]^{(1-L_2)} \times P_n^1(A_{0:1}, L_{0:1}). \end{aligned}$$

Again, it is easy to verify that the latter fluctuation model goes through P_n^1 at $\epsilon = 0$, with a score at $\epsilon = 0$ equal to $D_2^*(\cdot \mid P_n^1)$. Let $\epsilon_{n,2}$ denote the maximum likelihood estimate in the model, and let $P_n^2 = P_n^1(\epsilon_{n,2})$ be the corresponding update. Finally, let us introduce the third clever covariate for fluctuating the conditional distribution of L_1 under P_n^2, the "next" factor of $Q_n^2 = Q(\cdot \mid P_n^2)$:

$$\begin{aligned} H_{n,1}^* = {} & \frac{I(A_0 = 1)}{g(A_0 = 1 \mid X; P_n^2)} \times (1 - P_n^2(L_2 = 1 \mid A_{0:1} = 1_{0:1}, L_1 = 0, L_0) \\ & \times P_n^2(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = (0, 1), L_0) - P_n^2(L_2 = 0 \mid A_{0:1} = 1_{0:1}, L_1 = 0, L_0) \\ & \times P_n^2(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = 0_{1:2}, L_0)) \end{aligned}$$

and the related fluctuation model

$$\begin{aligned} P_n^2(\epsilon)(O) = {} & \prod_{j=2}^{3} Q(L_j \mid A_{0:j}, L_{0:j}; P_n^2) \\ & \times \prod_{j=1}^{2} g(A_j = 1 \mid X; P_n^2)^{A_j}(1 - g(A_j = 1 \mid X; P_n^2))^{1-A_j} \\ & \times \operatorname{expit}\Big(\operatorname{logit} Q(L_1 \mid A_0, L_0; P_n^2) + \epsilon H_{n,1}^*\Big)^{L_1} \\ & \times \Big[1 - \operatorname{expit}\Big(\operatorname{logit} Q(L_1 \mid A_0, L_0; P_n^2) + \epsilon H_{n,1}^*\Big)\Big]^{(1-L_1)} P_n^2(A_0, L_0). \end{aligned}$$

Once again, this fluctuation model goes through P_n^2 at $\epsilon = 0$, with its score at $\epsilon = 0$ equal to $D_1^*(\cdot \mid P_n^2)$. Let $\epsilon_{n,1}$ denote the maximum likelihood estimate in the model, and we define $P_n^* = P_n^2(\epsilon_{n,1})$. The first (and last) factor is the marginal distribution of L_0 under P_n^0, which is thus the empirical probability distribution of L_0. This is already a nonparametric maximum likelihood estimator, so that carrying out another updating step as above will result in an estimate of ϵ equal to zero. The TMLE of ψ_0 is the resulting substitution estimator $\Psi(P_n^*)$.

A closer look at the construction of P_n^* finally yields the following result:

Proposition 25.2. *It holds that*

(1) For each $1 \leq j \leq 3$, $D_j^*(\cdot \mid P_n^{3-j}) = D_j^*(\cdot \mid P_n^*)$.

(2) $Q(L_0; P_n^0) = \frac{1}{n}\sum_{i=1}^n I(L_0^{(i)} = L_0)$ *(expressed in words, the marginal distribution of* L_0 *is estimated with its empirical distribution), and* $P_n D^*(\cdot \mid P_n^*) = 0$.

(3) The TMLE of ψ_0, $\psi_n^* = \Psi(P_n^*)$, *satisfies*

$$\psi_n^* = \frac{1}{n}\sum_{i=1}^n \sum_{\ell_{1:2}\in\{0,1\}^2} Q(Y = 1 \mid A_{0:2} = 1_{0:2}, L_{1:2} = \ell_{1:2}, L_0^{(i)}; P_n^*)$$
$$\times Q(L_2 = \ell_2 \mid A_{0:1} = 1_{0:1}, L_1 = \ell_1, L_0^{(i)}; P_n^*) \times Q(L_1 = \ell_1 \mid A_0 = 1, L_0^{(i)}; P_n^*).$$

Item (1) in Proposition 25.2 is an example of the so-called *monotonicity* property of the clever covariates, which states that the clever covariate of the jth factor in Q_0 only depends on the future (later) factors of Q_0. This monotonicity property implies that the TMLE procedure presented above converges in one single step, referring to the iterative nature of the general TMLE procedure. A typical iterative TMLE procedure (Chap. 24 and Appendix A) would use the same logistic regression fluctuation models as presented above, but it would enforce a common ϵ across the different factors of Q_0, and thus updates all factors simultaneously at each maximum likelihood update step. This iterative TMLE converges very fast: in similar examples, experience shows that convergence is often achieved in two or three steps, and that most reduction occurs during the first step). Item (2) is of fundamental importance since it allows us to study the properties of $\Psi(P_n^*)$ from the point of view of the general theory of estimating equations. Item (3) just states that ψ_n^* is a plug-in estimator $\Psi(P_n^*)$ and provides a simple formula for evaluating ψ_n^*.

25.3.2 Merits of TMLE Procedure

Since the efficient influence curve $D^*(\cdot \mid P)$ is double robust and since P_n^* solves the efficient influence curve estimating equation, the general theory of estimating equations teaches us that the TMLE ψ_n^* enjoys remarkable asymptotic properties under certain assumptions. Stating the latter assumptions is outside the scope of this chapter. One often refers to such conditions as *regularity conditions*. Let $P_n^* = Q_n^* g_n^*$. The regularity conditions typically include the requirements that the sequence $(\psi_n^* : n = 1, \ldots)$ must belong to a compact set; that both Q_n^* and g_n^* must converge to some Q_1 and g_1 with at least one of these limits representing the truth; that the estimated efficient influence curve $D(\cdot \mid P_n^*)$ must belong to a P_0-Donsker class with P_0-probability tending to one; and that a second-order term that involves a product of $Q_n^* - Q_1$ and $g_n^* - g_1$ is $o_P(1/\sqrt{n})$. See Appendix A for more details.

The following classical result holds:

Proposition 25.3. *Under regularity conditions,*

(1) The TMLE ψ_n^* *consistently estimates* ψ_0 *as soon as either* Q_n^* *or* g_n^* *consistently estimates* Q_0 *or* g_0.

(2) If the TMLE consistently estimates ψ_0, then it is asymptotically linear: there exists D such that

$$\sqrt{n}(\psi_n^* - \psi_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} D(O^{(i)}) + o_P(1). \tag{25.5}$$

Equation (25.5) straightforwardly yields that $\sqrt{n}(\psi_n^* - \psi_0)$ is asymptotically Gaussian with mean zero and variance consistently estimated by

$$\frac{1}{n} \sum_{i=1}^{n} D_n(O^{(i)})^2,$$

where D_n is a consistent estimator of influence curve D.

If Q_n^* and g_n^* consistently estimate Q_0 and g_0 (hence P_n^* consistently estimates P_0), then $D = D^*(\cdot; P_0)$, so that the asymptotic variance is consistently estimated by

$$\frac{1}{n} \sum_{i=1}^{n} D^*(O^{(i)} \mid P_n^*)^2. \tag{25.6}$$

In this case, the TMLE is asymptotically efficient: its asymptotic variance is as small as possible (in the family of regular estimators).

Furthermore, if g_n^* is a maximum-likelihood-based consistent estimator of g_0, then (25.6) is a conservative estimator of the asymptotic variance of $\sqrt{n}(\psi_n^* - \psi_0)$ (it converges to an upper bound on the latter asymptotic variance).

Proposition 25.3 is the cornerstone of the TMLE methodology. It allows us to build confidence intervals. We assess how well such confidence intervals perform from a practical point of view through a simulation study in Sect. 25.4. For the estimation of the probability of success carried out in Sect. 25.5, we resort to the bootstrap to compute a confidence interval.

We emphasized how the TMLE benefits from advances in the theory of estimating equations. Yet, it enjoys some remarkable advantages over estimating equation methods. Let us briefly evoke the most striking in the context of this chapter:

(1) The TMLE is a substitution estimator. Thus, it automatically satisfies any constraint on the parameter of interest (here that the parameter of interest is a proportion and must therefore belong to the unit interval), and it respects the knowledge that the parameter of interest is a particular function of the data-generating distribution. On the contrary, solutions of an estimating equation may fail to satisfy such constraints.
(2) The TMLE methodology cares about the likelihood. The log-likelihood of the updated estimate of P_0, $\frac{1}{n} \sum_{i=1}^{n} \log P_n^*(O_i)$, is available, thereby allowing for the C-TMLE extension (Part VII).

25.3.3 Implementing TMLE

The TMLE procedure is implemented following the specification in Sect. 25.3.1. Only the details of the super learning procedure are missing. We chose to rely on a least squares loss functions, and on a collection of algorithms containing seven estimation procedures: generalized linear models, elastic net ($\alpha = 1$), elastic net ($\alpha = 0.5$), generalized additive models (degree = 2), generalized additive models (degree = 3), DSA, and random forest (ntree = 1000).

25.4 Simulations

The simulation scheme attempts to mimic the data-generating distribution of the DAIFI data set. We start with L_0 drawn from its empirical distribution based on the DAIFI data set and for $j = 0, \ldots, 2$, successively, $A_j \sim \text{Ber}(q_j(L_{0,1}, L_{0,2}, L_{0,3}))$ and $L_{j+1} \sim \text{Ber}(p_{j+1}(L_{0,1}, L_{0,2}, L_{0,3}))$, where for each $j = 0, 1, 2$, $p_{j+1}(L_{0,1}, L_{0,2}, L_{0,3}) = \text{expit}\,(\alpha_{1,L_{0,1}} + \alpha_{2,L_{0,1}} \log L_{0,2} + \alpha_{3,L_{0,1}} \log(5 + \min(L_{0,3}, 5)^5))$, and for each $j = 0, 1, 2$, $q_j(L_{0,1}, L_{0,2}, L_{0,3}) = \text{expit}\,(\beta_{1,L_{0,1}} + \beta_{2,L_{0,1}} L_{0,3})$. The values of the α- and β-parameters are reported in Table 25.1.

Regarding the empirical distribution of L_0, the IVF unit random variable $L_{0,1}$ follows a Bernoulli distribution with parameter approximately equal to 0.517. Both conditional distributions of age $L_{0,2}$ given the IVF unit are Gaussian-like, with means and standard deviations roughly equal to 33 and 4.4. The marginal distribution of the random number $L_{0,3}$ of embryos transferred or frozen has mean and variance approximately equal to 3.3 and 7.5, with values ranging between 0 and 23 (only 20% of the observed $L_{0,3}^{(i)}$ are larger than 5). We refer to Table 25.3 in Sect. 25.5 for a comparison of the empirical probabilities that $A_j = 1$ and $L_j = 1$ computed under the empirical distribution of a simulated data set with 10,000 observations and the empirical distribution of the DAIFI data set.

The super learning library is correctly specified for the estimation of the censoring mechanism, and misspecified for the estimation of the Q-factor. Indeed, $L_{0,2}$ plays a role in $p_{j+1}(L_{0,1}, L_{0,2}, L_{0,3})$ through its logarithm, and $L_{0,3}$ through $\log(5 + \min(L_{0,3}, 5)^5)$. We choose this expression because $x \mapsto \log(5 + \min(x, 5)^5)$ cannot be well approximated by a polynomial in x over $[0, 23]$. Furthermore, the true

Table 25.1 Values of the α- and β-parameters used in the simulation scheme

	IVF cycle j			
Parameters	0	1	2	3
$100 \times \alpha_{\cdot,0}$	–	(61, −55, 4.5)	(13, −45, 1.2)	(60, −40, 1.5)
$100 \times \alpha_{\cdot,1}$	–	(65, −70, 3.3)	(19, −49, 1.7)	(80, −50, 1)
$100 \times \beta_{\cdot,0}$	(40, 10)	(−45, 32)	(−30, 5)	–
$100 \times \beta_{\cdot,1}$	(40, 9)	(−50, 34)	(−40, 6)	–

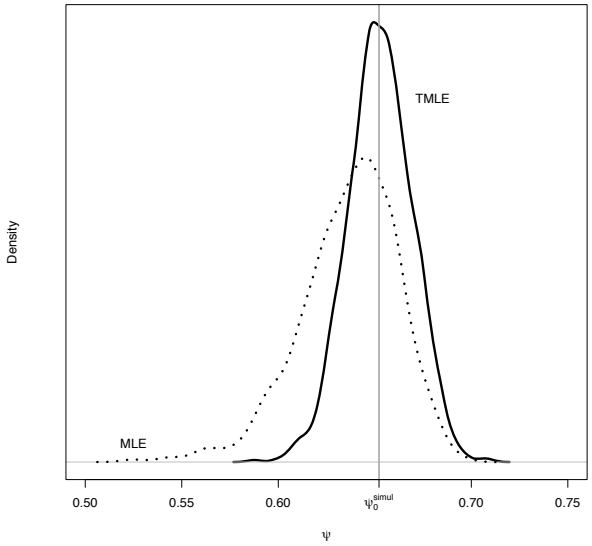

Fig. 25.1 MLE and TMLE empirical densities

Table 25.2 Simulation results

	Empirical		
Estimator	Bias	MSE	Cover (p-value)
MLE	−0.01571	0.00065	–
TMLE	0.00052	0.00027	0.958 (89%)

value of the parameter of interest for this simulation scheme can be estimated with great precision by Monte Carlo. Using a simulated data set of one million observations under the intervention (1,1,1) yields $\psi_0^{\text{simul}} = 0.652187$, with a 95% confidence interval equal to $(0.6512535, 0.6531205)$.

We repeat $B = 1000$ times the following steps: (1) simulate a data set with sample size $n = 3001$ according to the simulation scheme presented above and (2) estimate ψ_0^{simul} with $\Psi(P_{n,b}^*) = \psi_{n,b}^*$, the bth TMLE based on this bth data set. In order to shed some light one the properties of the TMLE procedure, we also keep track of the initial maximum-likelihood-based substitution estimator $\Psi(P_{n,b}^0)$, based on the bth data set.

We summarize the results of the simulation study in Table 25.2. They are illuminating: the MLE is biased, whereas the TMLE is unbiased. In the process of stretching the initial MLE into the updated TMLE in the direction of the parameter of interest, the update step not only corrects the bias but also diminishes the variance. Those key features are well illustrated in Fig. 25.1, where it is also seen that the TMLE is approximately normally distributed.

Let us now investigate the validity of the coverage guaranteed by the 95% confidence intervals based on the central limit theorem satisfied by ψ_n^*, using (25.6) as

an estimate of the asymptotic variance. Since the super learner g_n^* is a consistent estimator of g_0, the latter estimate of the asymptotic variance of the TMLE is sensible, and may be slightly conservative due to the misspecification of Q_n^*. Among the $B = 1000$ 95% confidence intervals, 958 contain the true value ψ_0^{simul}. This is strongly in favor of the conclusion that the confidence intervals do meet their requirement. The probability for a binomial random variable with parameter $(B, 95\%)$ to be less than 958 equals 89%.

25.5 Data Application

We observe $n = 3001$ experimental units. We report in Table 25.3 the empirical probabilities of $A_j = 1$ (each $j = 0, 1, 2$) and $L_j = 1$ ($j = 0, 1, 2, 3$) for all IVF cycles. It is obvious from these numbers that the censoring mechanism plays a great role in the data-generating experiment. We applied the TMLE methodology and obtained a point estimate of ψ_0 equal to $\psi_n^* = 0.505$. The corresponding 95% confidence interval based on the central limit theorem, using (25.6) as an estimate of the asymptotic variance of $\sqrt{n}(\psi_n^* - \psi_0)$, is equal to $(0.480, 0.530)$.

However, we have no certainty of the convergence of g_n^* to g_0 (which would guarantee that the confidence interval is conservative). Therefore we also carried out a bootstrap study. Specifically, we iterated $B = 1000$ times the following procedure. First, draw a data set of $n = 3001$ observations from the empirical measure P_n; second, compute and store the TMLE $\psi_{n,b}^*$ obtained on this data set. This results in the following 95% confidence interval (using the original ψ_n^* as center of the interval): $(0.470, 0.540)$, which is wider than the previous one.

As a side remark, the MLE $\Psi(P_n^0)$ updated during the second step of the TMLE procedure (applied to the original data set) is equal to 0.490. We also note that the TMLE ψ_n^* falls between the estimates obtained in Soullier et al. (2008) by multiple imputation and the Kaplan–Meier method. The probability of success of a program of at most four IVF cycles may be slightly larger than previously thought. In conclusion, future participants in a program of at most four IVF cycles can be informed that approximately half of them may subsequently succeed in having a child.

Table 25.3 Empirical probabilities that $A_j = 1$ and $L_j = 1$ based on a simulated data set of 10,000 observations and the DAIFI data set

	Simulated data set		DAIFI data set	
	Empirical probability of		Empirical probability of	
IVF cycle j	$A_j = 1$	$L_j = 1$	$A_j = 1$	$L_j = 1$
0	73%	21%	75%	22%
1	57%	32%	59%	32%
2	46%	37%	49%	35%
3	–	40%	–	37%

25.6 Discussion

We studied the performance of IVF programs and provided the community with an accurate estimator of the probability of delivery during the course of a program of at most four IVF cycles in France (abbreviated to probability of success). We first expressed the parameter of interest as a functional $\Psi(P_0)$ of the data-generating distribution P_0 of the observed longitudinal data structure $O = (L_0, A_0, L_1, A_1, L_2, A_2, Y)$. Subsequently, we applied the TMLE. Under regularity conditions, the estimator is consistent as soon as at least one of two fundamental components of P_0 is consistently estimated; moreover, the central limit theorem allowed us to construct a confidence interval. These theoretical properties are illustrated by a simulation study. We obtained a point estimate that is approximately equal to 50%, with a 95% confidence interval given by (48%, 53%). Earlier results obtained by Soullier et al. (2008) based on the multiple-imputation methodology were slightly more pessimistic, with an estimated probability of success equal to 46% and (44%, 48%) as 95% confidence interval.

These authors also considered another approach that involves phrasing the problem of interest as the estimation of a survival function based on right-censored data. The key to this second approach is that the probability of success coincides with the probability $P(T \leq 3)$, where T is the number of IVF cycles attempted after the first one till the first successful delivery. Our observed longitudinal data structure O is equivalent to the right-censored data structure $O' = (W, \min(T, C), I(T \leq C))$, where $W = L_0$ and $C = \min(0 \leq j \leq 3 : A_j = 0)$ with the additional convention $A_3 = 0$. Neglecting the baseline covariates W and assuming that the dropout time C is independent of T, Soullier et al. (2008) estimated the probability of success by $1 - S_n(3)$, S_n being the Kaplan–Meier estimate of the survival function of T. Although the TMLE methodology to address the estimation of $P(T \leq 3)$ (incorporating the baseline covariates) is well understood (Chaps. 17 and 18), we choose to adopt the point of view of a longitudinal data structure rather than that of a right-censored data structure. From the survival analysis point of view, our contribution is to incorporate the baseline covariates in order to improve efficiency and to allow for informative censoring. We finally emphasize that an extension of the TMLE procedure presented in this chapter will allow, in future work, to take into account the successive number of embryos transferred or frozen at each IVF cycle (instead of the sole number at the first IVF cycle), thereby acknowledging the possibility that this time-dependent covariate may yield time-dependent confounding.

Acknowledgements

The author would like to thank E. de la Rochebrochard, J. Bouyer (Ined; Inserm, CESP; Univ Paris-Sud, UMRS 1018), and S. Enjalric-Ancelet (AgroParisTech, Unité Mét@risk) for introducing him to this problem, as well as the Cochin and Clermont-Ferrand IVF units for sharing the DAIFI data set.