under treatment randomization as in successfully conducted experiments. This allows identifying the ATE: $\Delta_1 = E[Y_1|D=1] - E[Y_1|D=0]$. Furthermore, we assume the mediator to be weakly monotonic in the treatment.

Assumption 8: Weak monotonicity of the mediator in the treatment.

$\Pr(M(1) \geq M(0)) = 1$.

Assumption 8 is standard in the instrumental variable literature on local average treatment effects when denoting by D the instrument and by M the endogenous regressor, see and . It rules out the existence of defiers. As discussed in the Appendix , the total ATE $\Delta_1 = E[Y_1|D=1] - E[Y_1|D=0]$ and QTE $\Delta_1(q) = F^{-1}_{Y_1|D=1}(q) - F^{-1}_{Y_1|D=0}(q)$ for the entire population are identified under Assumption 7. Furthermore, Assumptions 7 and 8 yield the strata proportions, denoted by $p_\tau = \Pr(\tau)$, as functions of the conditional mediator probabilities given the treatment, which we denote by $p_{(m|d)} = \Pr(M=m|D=d)$ for $d, m \in \{0,1\}$ (see Appendix):

$$p_a = p_{1|0}, p_c = p_{1|1} - p_{1|0} = p_{0|0} - p_{0|1}, p_n = p_{0|1}. \tag{1}$$

Furthermore, Assumptions 2, 7, and 8 imply that (see Appendix)

$$\Delta_{0,c} = E[Y_0(1,1) - Y_0(0,0)|c] = \frac{E[Y_0|D=1] - E[Y_0|D=0]}{p_{1|1} - p_{1|0}} = 0. \tag{2}$$

Therefore, a rejection of the testable implication $E[Y_0|D=1] - E[Y_0|D=0] = 0$ in the data would point to a violation of these assumptions. Assumptions 7 and 8 permit identifying additional parameters, namely the total, direct, and indirect effects on compliers, and the direct effects on never- and always-takers, as shown in Theorems 3 to 5. This follows from the fact that defiers are ruled out and that the proportions and potential outcome distributions of the various principal strata are not selective w.r.t. the treatment.

Theorem 3: Under Assumptions 1–3, 7-8,