**Highlights**

* Schemes evaluated based on processes (compared to outcomes) appear to have larger effects
* Larger incentives are likely to have greater effects than smaller incentives
* Better evaluated P4P schemes show smaller estimates of effects