Marginal structural models: much ado about (almost) nothing - Part 2 -

In 2011 we published a critique of inverse probability of treatment weighting (IPTW),[1] a method to remove confounding bias.[2] First, we showed (like others) that the IPTW estimator may be viewed as an effect estimator in a "target population" of exposed and unexposed who are matched on the confounder (L). Then, we emphasized that just like standardization, the unconfounded association from a marginal structural model depends on the choice of an arbitrary "target population" and does not provide any conjecture about the effect in any human being.

We wrote:

"Attempts to justify one target over another switch, incoherently, from causal claims about causal parameters of finite populations to causal claim about individual patients. For example, the subsets $L = 1$ and $L = 0$ are of interest, because chances are that the treatment will benefit Jim ($L = 1$) more than it will benefit Tim ($L = 0$). But without axioms, the stated definitions behind a marginal structural model cannot form the basis of the claim. There is no inference from the causal parameter of an arbitrary target to Jim or to Tim, because they may be placed in infinitely many targets, together or separately. You cannot play the game in both fields. If causal parameters are defined as counterfactuals in finite populations, there is no inference from any study about the effect in any patient within, or outside, the study. Throw determinism to the mix [14], and science offers no conjectural inference on any patient, no matter which model is fit. You may just as well make a treatment decision by tossing a coin, the outcome of which was predetermined, too . . ."[1]

Is the IPTW estimator therefore useless? Not necessarily. Without invoking "fictional observations in a target population", we showed an alternative view of the estimator: a weighted average of estimated effects in strata of the confounder.

"Under our axiomatization, marginal structural models estimate a weighted average of effects, and do not present a new idea in the domain of deconfounding. Whenever a weighted average and the weighting scheme can both be rationalized, the models are acceptable. Whenever a weighted average does not estimate an effect (i.e. important effect modification is ignored) or the weights are arbitrary, the models should not be fit."[1]

Finally, we argued that

"…every method for deconfounding, except randomization and actual matching, requires the blocking of a confounding path by conditioning – even when it claims to be doing something else."[1]

Not much later, Weuve et al. published an article on another version of marginal structural models: inverse probability of attrition weighting (IPAW).[3] First, that's a perpetuated misnomer.[4] The method should be called "inverse probability of non-attrition weighting" (IPNAW), as any reader can tell. Second, this time the marginal structural model helps to remove colliding bias.

Thought bias

Before discussing the method, we can't avoid commenting on the subject matter of the article: smoking and cognitive decline.

Unaware of thought bias[5-7] (or denying its existence), Weuve and her colleagues believe that "cognitive decline"—the output of a function of two variables—is a causal entity in the natural world, which has causes and effects. Weuve et al. are not alone in failing to appreciate the absurdity of that thought. The number of the causes of $f(V_1,V_2)=V_2-V_1$ is as large as the number of the causes of $f(V_1,V_2)=V_2/V_1$, $f(V_1,V_2)=(V_2-V_1)/V_1$, $f(V_1,V_2)=V_1 \times V_2$, and $f(V_1,V_2,U_3)=(V_2-V_1/U_3)^e$, namely zero.

Derived variables such as V_2-V_1 have neither causes nor effects: they are mathematical entities (objects of World 3),[8] rather than natural properties of objects. Indeed, smoking at one time might affect cognition at another time, but it has no effect (not "null effect") on the subtraction (or division or multiplication) of the cognition at two times. The motivating causal diagram of Weuve at al. does not exist *a priori* in causal reality. Nor does the causal parameter they claim to estimate. Anyone who thinks differently should find the courage to explain a counter viewpoint in the public domain.

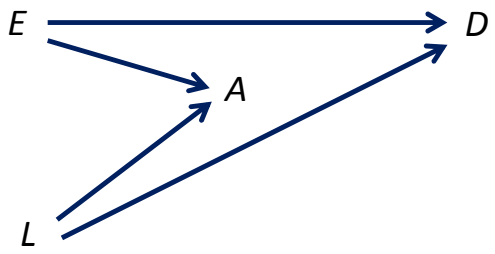
An example of colliding bias

Weuve et al. chose to present their new method in the context of a set of other unnecessary ideas: regression models, waves of follow-up, time-varying variables, multiple causes of a collider, cross-products, functional forms of time, and more. We think it was a poor pedagogical choice. A new method should be presented in the simplest possible form, without mixing it with other methodological challenges. To focus on the core of the matter, we

will reduce the problem to merely four variables and three time points. All that IPNAW claims to be doing can be examined in this simple example.

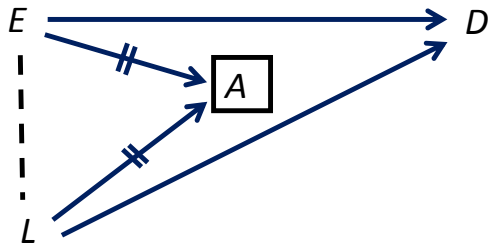
Figure 1 shows a causal structure: *E* ("exposure") is smoking status; *D* ("disease") is cognition; *A* is interim vital status (a type of attrition); and *L* is some cause of both vital status and cognition. The variable *A* is a collider on the path *E*→*A*←*L*→*D*, whereas *E* and *L* are *colliding variables* (through their colliding arrows). We assume that all variables are binary, except cognition.

Figure 1. A causal structure



In a simple-minded analysis, we condition on *A* (restrict the sample to *A*=1), because the value of *D* is set to be missing in those who have died in the interim (*A*=0). If *E* and *L* are effect modifiers with respect to *A* (on the probability ratio scale), we observe a structure of colliding bias (Figure 2).[7] *E* and *L* are now associated (denoted by a dashed line), and the path *E*--*L*→*D* contributes to the conditional association between *E* and *D*.

Figure 2. Colliding bias



Two basic solutions may be proposed:

Condition on *L*, an intermediary variable on the path of colliding bias, which will eliminate the induced path (Figure 3).

Avoid conditioning on *A* by one of the following methods:

- Assign a value of *D* to deceased people. For example, the cognition of dead people may be zero (or some value you would give to "brain death".)

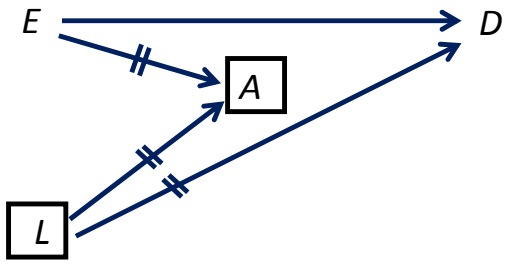
- Impute the missing values of *D* by some method. In particular, the missing values may be imputed from variables that are associated with *D*, such as *L*. Notice the trade-off: colliding bias is traded for information bias. No free lunch in science.

Again, we previously wrote that

"…every method for deconfounding, except randomization and actual matching, requires the blocking of a confounding path by conditioning – even when it claims to be doing something else."[1]

We now add: every method that removes colliding bias either trades colliding bias for information bias, or requires the blocking of a path of colliding bias by conditioning—even when it claims to be doing something else.

Figure 3. Colliding bias removed



Inverse probability of non-attrition weighting

At first glance, IPNAW contains some features of imputation, but the full story will be revealed later. In our example, the estimate is derived as follows.

- Compute the probability of being alive (*A*=1) from *E* and *L*. In our simple case of binary variables, you can compute these probabilities by hand. There are four:

  Pr (*A*=1|*E*=1, *L*=1)
  Pr (*A*=1|*E*=1, *L*=0)
  Pr (*A*=1|*E*=0, *L*=1)
  Pr (*A*=1|*E*=0, *L*=0)

- Weigh each observation of an alive person according to the inverse of his/her probability above. For example, if Smith survived (*A*=1) and

his other values were E=1, L=1, D=122, the weight or his data would be 1/Pr (A=1|E=1, L=1).

- Compute the association between E and D, using the observations of alive people and their weights. This is the IPNAW estimate.

The meaning of weighting observations

Weighting observations adds fictional observations to the data set. If Smith's probability of surviving was 0.25 (weight = 4), his row of data (A=1, E=1, L=1, D=122) is effectively replicated three more times. (You can technically create a new data set by copying and pasting his data three times, assigning a unique identifier to each row of data and repeating the procedure for other observations.)

Oddly enough, Weuve and her colleagues claim that "adding fictional observations" is just a metaphor for weighting observations, and the appropriate statistical interpretation of inverse probability weighted estimation has to do with ensuring statistical independence between variables.[9] They undoubtedly forgot the usual order in statistics: first—values; then—statistical relations between the variables that contain them. The variables in question are fictional variables that result from adding fictional observations; we cannot speak about independence between these variables until the fictional observations are added. Therefore, adding fictional observations is the essence of the matter, not a metaphor. Should they redefine "variable" to mean "something that contains values and their weights", we will let them count the followers.

Features of the fictional data set

It is easy to see that the number of fictional observations that are added by IPNAW is identical to the number of dead people in each stratum of L and E (Appendix). Therefore, the original sample size is maintained. Since no value of D is missing, all the "observations" can now be used to estimate the effect of E on D. Did we avoid conditioning on A? No, we did not. All the observations, real and fictional, still take the value A=1. (Remember the copying and pasting of the data of alive people?)

Still, colliding bias was removed. How come?

Colliding bias was removed *because E and L are not associated*, despite continued conditioning on A. Notice that the marginal association between E and L in the fictional data set is identical to their marginal association in the real data set (Appendix). Since E and L should have been independent *before* any analysis (based on Figure 1), they are also independent in the fictional data. Stated differently, exposed and unexposed are matched on L in the fictional data set.

The similarity to IPTW is obvious. In the fictional data set of IPTW, confounding bias is absent because exposed and unexposed are matched on the confounder.[1] In the fictional data set of IPNAW, colliding bias is absent because exposed and unexposed are matched on the colliding variable, L.

How was the independence restored when switching from analysis of the real data (Figure 2) to analysis of the fictional data? How did we succeed to erase the dashed line between E and L despite continued conditioning on A? As we mentioned earlier, there are only two methods to dissociate E from L after conditioning on A: Conditioning on L (Figure 3) or imputing the values of D for A=0. While it may seem that we are imputing the values of D by introducing fictional data, we will show (next section) that *we are actually conditioning on L and taking a weighted average of the effect of E on D across the strata of L*.

Just like IPTW which blocks a confounding path by conditioning on an intermediary variable (the confounder),[1] IPNAW blocks a path of colliding bias by conditioning on an intermediary variable (the colliding variable). Similar to IPTW which trades confounding bias for increased variance,[1] IPNAW trades colliding bias for increased variance.

The IPNAW estimator: a weighted average of stratum-specific effects

We showed that the IPTW estimator may be written as a weighted average of the effect of E on D in the two strata of the confounder.[1] Likewise, the IPNAW estimator may be written as a weighted average of the effect of E on D in the two strata of the colliding variable.

Using the Appendix, we show the proof for the probability ratio of a binary outcome. (Similar proofs can be developed for other kinds of outcomes.) Notice first that the four original event probabilities in the strata of E and L ("proportion of observed events") are maintained in the fictional data. In other words, the original proportions were exported by IPNAW to the fictional data, as was the case in IPTW.[1]

Let P_E and P_U, denote, respectively, the proportion of fictional events in exposed and unexposed in the fictional population. We will show that the proportion ratio, P_E/P_U ("the marginal association"), can be written as a weighted average of the L-specific proportion ratios in the real data (the post-attrition data set). That is,

$$P_E/P_U = (P_{L0,E1}/P_{L0,E0})w_{L0} + (P_{L1,E1}/P_{L1,E0})w_{L1}$$

where $w_{L0} + w_{L1} = 1$

Proof: (see notation in the Appendix)

$$P_E = (P_{L0,E1}\, x\, n_{L0,E1} + P_{L1,E1}\, x\, n_{L1,E1})/T_E$$

$$P_U = (P_{L0,E0}\, x\, n_{L0,E0} + P_{L1,E0}\, x\, n_{L1,E0})/T_U$$

Let

$w_1 = n_{L0,E1}/T_E$, the proportion of *L*=0 in exposed
$w_2 = n_{L1,E1}/T_E$, the proportion of *L*=1 in exposed
$w_3 = n_{L0,E0}/T_U$, the proportion of *L*=0 in unexposed
$w_4 = n_{L1,E0}/T_U$, the proportion of *L*=1 in unexposed

Then,

$$P_E = P_{L0,E1}\, x\, w_1 + P_{L1,E1}\, x\, w_2$$

$$P_U = P_{L0,E0}\, x\, w_3 + P_{L1,E0}\, x\, w_4$$

$$P_E/P_U = \frac{P_{L0,E1}\, x\, w_1 + P_{L1,E1}\, x\, w_2}{P_{L0,E0}\, x\, w_3 + P_{L1,E0}\, x\, w_4}$$

Recall that *E* and *L* should not be associated in the real data (Figure 1). Therefore,

$w_3 = w_1$
$w_4 = w_2$

and

$$P_E/P_U = \frac{P_{L0,E1}\, x\, w_1 + P_{L1,E1}\, x\, w_2}{P_{L0,E0}\, x\, w_1 + P_{L1,E0}\, x\, w_2} =$$

$$= \frac{\frac{P_{L0,E1}}{P_{L0,E0}} x P_{L0,E0}\, x\, w_1 + \frac{P_{L1,E1}}{P_{L1,E0}} x P_{L1,E0}\, x\, w_2}{P_{L0,E0}\, x\, w_1 + P_{L1,E0}\, x\, w_2}$$

Let

$$w_{L0} = \frac{P_{L0,E0}\, x\, w_1}{P_{L0,E0}\, x\, w_1 + P_{L1,E0}\, x\, w_2}$$

$$w_{L1} = \frac{P_{L1,E0}\, x\, w_2}{P_{L0,E0}\, x\, w_1 + P_{L1,E0}\, x\, w_2}$$

Then, $w_{L0} + w_{L1} = 1$

and

$$P_E/P_U = (P_{L0,E1}/P_{L0,E0})w_{L0} + (P_{L1,E1}/P_{L1,E0})w_{L1}$$

In retrospect, no surprise was in store because the theorems of a causal diagram speak in a clear language. If Figure 1 depicts the relevant part of causal reality, and you are forced to condition on *A* because *D* is missing when *A*=0 (Figure 2), there is only one way to block the induced path *E*--*L*→*D*: conditioning on *L* as well (Figure 3). Any impostor that claims to achieve that goal differently (such as the IPNAW estimator) must have included conditioning of the association between *E* and *D* on the variable *L*.

As with any weighted average, two questions should be asked: Do we prefer a weighted average across the strata of *L* over *L*-specific estimates? If so, can the weights be rationalized? The very same questions were raised and explored in IPTW.[1] Not surprisingly, the answers are similar, so they are summarized here briefly.

- If *E* and *L* are effect modifiers with respect to *D*, and effect modification is strong, a weighted average embeds substantial effect modification bias.[7] Neither the IPNAW estimator nor any other *L*-weighted estimator should be computed.

- Depending on the measure of effect, the weights may be rationalized in some cases, but not others. The burden of justification rests on the user.

Why not use "old" methods for conditioning on *L*?

Indeed, why not? What does weighting by the IPNAW estimator add that is missing from classic weighting by other methods, such as the Mantel-Haenszel estimators and regression?

As with IPTW, Weuve et al. enter the muddy territory of partial effects, direct effects, and the effect of time-dependent exposures—problems that do not exist under at least one axiomatization of causality.[1] We are still waiting to read another axiomatization that allows these problems to survive. Make no mistake: we are looking for formal axioms about the way causality *works*—not the usual mix of methodology and definitionology.

The "target population" (again)

IPTW estimates the effect in one "target population" but the principles can be used for other targets.[10] The same is true for IPNAW. You may compute an IPNAW estimate for different populations by using a different set of weights provided they satisfy w_3=w_1 and w_4=w_2. Here is an example:

$w_1 = 7/ \Pr(A=1|E=1, L=0)$
$w_2 = 9/ \Pr(A=1|E=1, L=1)$
$w_3 = 7/ \Pr(A=1|E=0, L=0)$
$w_4 = 9/ \Pr(A=1|E=0, L=1)$

where the proportions of fictional evens in exposed and unexposed are given by the following formulae:

$$P_E = P_{L0,E1}\, x\, w_1 + P_{L1,E1}\, x\, w_2$$

$$P_U = P_{L0,E0}\, x\, w_3 + P_{L1,E0}\, x\, w_4$$

In other words, you may estimate as many effects as you wish (for one cause-and-effect pair). Then, you can use the estimates to articulate "science-based policies"—claims to people about effects in "targets", without telling them two crucial pieces of information: 1) Each person may be placed in many targets; 2) The claims don't apply to any specific person.[11] The uninformed public will probably follow the "policy", but what about scientists?

Statistical tools are not tailored only to samples of human beings. In principle, a marginal structural model may be fit to estimate an effect in a finite target of mice or of muons. Those who may do so should explain to biologists and physicists why this method is helpful in causal inquiry. Policy making for mice? Policy making for muons?

The deterministic debate

Chaix et al. wrote a commentary on the article by Weuve et al.,[12] and some of the authors replied (Weuve listed last).[9] The exchange centers on undefinability of the outcome in people who died, on counterfactuals, and especially on the validity of IPNAW under "principal stratification".

Chaix et al. explain:

"the so-called principal-stratification approach invites us to distinguish 3 different "principal strata" in this study: (i) people who would survive until outcome measurement, whether they smoke or not ("always-survivors"); (ii) those who would die before measurement, whatever their smoking status; and (iii) those who would survive as nonsmokers but would die as smokers."[12]

Those who are not familiar with the principle of "principal stratification" may recall the old classification into four types of people, under determinism:[13] "immune", "doomed", "causative", and "preventive". (The "preventive" type is not allowed to exist here.)

Chaix et al. question the appropriateness of IPNAW under principal stratification, and Weuve and her colleagues prove that IPNAW does provide average deterministic effects in the target. That is, the estimator corresponds to math that incorporates the deterministic types, which Chaix et al. mention. Before celebrating the victory, let's recall the merit of an average effect on some disease—under determinism. How about a null effect of the exposure when each of the four types makes up a quarter of some finite target (of course, unbeknown to us)?[13] Let's form a policy: "To all members of the target (call to check your membership status): the exposure may be ignored."

Definability

Chaix et al. question the "replacing [of] dead participants by cloning the living".[12] They draw a distinction between a "real target population for which the outcome is defined" and a "pseudo-population that has no real existence" (because the outcome is undefined in dead people).

"Undefined" is another misnomer.[9] "To define" means to replace a phrase whose meaning is already clear (called definiens) with another, shorter phrase or word (called definiendum).[14] So what does "undefined" mean? Inability to make a linguistic substitution?

Contrary to prevailing misconception,[15] causal reality and the studying of causal reality are not a matter of definitions.[1] The argument of Chaix et al. should read: "The outcome variable does not exist in dead people". Unlike a derived variable ("cognitive decline") which never exists in causal reality, a natural variable exists as long as the object to which it is attached continues to exist. But when do a person and his properties cease to exist? For example, does brain mass no longer exist five minutes after the death was pronounced? Does cognition cease to exist upon death, or does it take the value "no function" for some time? That's a matter for a serious debate about the nature of the object, rather than a discussion of definability.

Just a matter of information bias?

Information bias takes one of two forms: wrong values (type 1) and no values (missingness, type 2). According to the traditional view of marginal structural models, confounding bias and colliding bias are problems of missing data in a target population,[16] which may be solved by filling in the missing values. Both types of bias are subsumed under the second type of information bias.

Read the last sentence again and you might appreciate the sweeping nature of the

epistemological claim. Confounding bias and colliding bias are not inherent in the difficulty of learning about a natural feature (causation) from a mathematical tool (association). They are explained by not having "the right data". Had we just had the right data for a finite target population (real or imaginary)—the problems would have disappeared. And if we don't have the right data, we can solve both problems by making up values of variables for real or imaginary members of the target.

If this viewpoint were correct, IPTW and IPNAW should have replaced type 2 information bias (missingness) with type 1 information bias (wrong values): the missing values were eliminated, but the imputed values are not necessarily the correct values. Our analysis shows, however, that the biases in question have been removed—without introducing new bias—because IPTW and IPNAW are equivalent to the conditioning that is sufficient to remove confounding bias and colliding bias, respectively. Therefore, confounding bias and colliding bias are two distinct categories of bias.[7] No, they are not just a matter of missing data (information bias).

References

1. Shahar E, Shahar DJ: Marginal structural models: much ado about (almost) nothing. *Journal of Evaluation in Clinical Practice* 2013;19:214-22 [Epub Aug 23, 2011]
2. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11:550-560
3. Weuve J, Tchetgen Tchetgen EJ, Glymour MM, Beck TL, Aggarwal NT, Wilson RS, Evans DA, Mendes de Leon CF. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology* 2012;23:119-28
4. Gottesman RF, Rawlings AM, Sharrett AR, Albert M, Alonso A, et al. Impact of differential attrition on the association of education with cognitive change over 20 years of follow-up: the ARIC neurocognitive study. *American Journal of Epidemiology* 2014;179:956-66
5. Shahar E, Shahar DJ: Causal diagrams, information bias, and thought bias. *Pragmatic and Observational Research* 2010;1:33–47
6. Shahar E, Shahar DJ: Causal diagrams and change variables. *Journal of Evaluation in Clinical Practice* 2012;18:143–148
7. Shahar E, Shahar DJ. Causal diagrams and three pairs of biases. In: *Epidemiology – Current Perspectives on Research and Practice* (Lunet N, Editor). 2012:pp 31-62 www.intechopen.com/books/epidemiology-current-perspectives-on-research-and-practice
8. Popper KR. The open universe: an argument for indeterminism. London, Routledge; 1988:113-130.
9. Tchetgen Tchetgen EJ, Glymour MM, Shpitser I, Weuve J. To weight or not to weight? On the relation between inverse-probability weighting and principal stratification for truncation by death. *Epidemiology* 2012;23:132-7
10. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology* 2003;14:680–686
11. Shahar E. Estimating causal parameters without target populations. *Journal of Evaluation in Clinical Practice* 2007;13:814-816
12. Chaix B, Evans D, Merlo J, Suzuki E. Weighing up the dead and missing: reflections on inverse-probability weighting and principal stratification to address truncation by death. *Epidemiology* 2012;23:129–131
13. Maldonado G, Greenland S. Estimating causal effects. *International Journal of Epidemiology* 2002;31:422-9
14. Shahar E, Shahar DJ. Causal diagrams, the placebo effect, and the expectation effect. *International Journal of General Medicine* 2013;6:821-8
15. Hernan MA. A definition of causal effect for epidemiological research. *Journal of Epidemiology and Community Health* 2004;58:265-271
16. Platt RW, Delaney JA, Suissa S. The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding. *European Journal of Epidemiology* 2012;27:77-83

Appendix: A simple example of inverse probability of non-attrition weighting

	Initial number of observations (n)	Final number of observations post attrition (n*)	Number of observed events (e)	Proportion of observed events	Weight (w) = $1/Pr(A=1\|L,E)$	Fictional number of observations (n* x w)	Fictional number of events (e x w)	Proportion of fictional events
***E*=1**								
$L=0$	$n_{L0,E1}$	$n^*_{L0,E1}$	a	$P_{L0,E1}=a/n^*_{L0,E1}$	$n_{L0,E1}/n^*_{L0,E1}$	$n_{L0,E1}$	$a(n_{L0,E1}/n^*_{L0,E1})$	$P_{L0,E1}=a/n^*_{L0,E1}$
$L=1$	$n_{L1,E1}$	$n^*_{L1,E1}$	b	$P_{L1,E1}=b/n^*_{L1,E1}$	$n_{L1,E1}/n^*_{L1,E1}$	$n_{L1,E1}$	$b(n_{L1,E1}/n^*_{L1,E1})$	$P_{L1,E1}=b/n^*_{L1,E1}$
Total	T_E					T_E		
***E*=0**								
$L=0$	$n_{L0,E0}$	$n^*_{L0,E0}$	c	$P_{L0,E0}=c/n^*_{L0,E0}$	$n_{L0,E0}/n^*_{L0,E0}$	$n_{L0,E0}$	$c(n_{L0,E0}/n^*_{L0,E0})$	$P_{L0,E0}=c/n^*_{L0,E0}$
$L=1$	$n_{L1,E0}$	$n^*_{L1,E0}$	d	$P_{L1,E0}=d/n^*_{L1,E0}$	$n_{L1,E0}/n^*_{L1,E0}$	$n_{L1,E0}$	$d(n_{L1,E0}/n^*_{L1,E0})$	$P_{L1,E0}=d/n^*_{L1,E0}$
Total	T_U					T_U		