Statistical decision theory with economic incentives

Aleksey Tetenov (University of Bristol)

Cemmap masterclass
Statistical decision theory for treatment choice and prediction
May 30-31, 2017

Motivation:

Pharmaceutical companies seek approval of their new drugs (so they could profit from them). To convince the regulator, they commission costly clinical trials that yield credible but imprecise statistical evidence (analyzed by hypothesis testing).

Researchers try to gain acceptance of their theories (from which they will benefit) by undertaking costly data collection or analysis (also analyzed by hypothesis testing).

Conventional statistical/econometric practice:

Null hypothesis testing: accept H_1 in a way that controls test size

$$P(\text{Type I error}|H_0) < 5\%$$

Hypothesis tests of $H_0 : \theta \leq 0$ (ineffective treatment) are used for treatment choice when it is framed as a binary choice between implementing an innovation and the status quo
- Explicit in international guidelines for drug approval.
- Implicit everywhere (from submission/publication decisions in scientific journals to newspaper articles).

Conventional test levels are arbitrary.

Widely criticized across many fields, but lives on.

Source: Tetenov (2016), "An economic theory of statistical testing," Cemmap working paper CWP50/16

Frame the statistical testing procedure as a strategy in a game against self-interested and informed proponents, rather than a game against nature.

Shows an environment in which classical null hypothesis testing criterion is rational

Derives a problem-specific test level α (not based on convention)

Main ideas

Null hypothesis testing is a minimax strategy for the regulator. It is reasonable if there **could be** lots of bad proposals.

Sufficiently low probability of approval (test size) deters the proponent from collecting statistical evidence in a costly and risky trial.

As a result, "null hypotheses" do not get tested.

The statistical procedure is designed to be a deterrent, whose strength depends on the true state of the world.

Its aim is NOT to infer the state of the world from the data, but to provide incentives for potential proponents to act on their information about it.

What is so strange about hypothesis testing?

Textbook way to motivate one-sided test of $H_0 : \theta \leq 0$ vs $H_1 : \theta > 0$ by "statistical decision theory:"

Two actions: accept H_1 or accept H_0.
Loss function: lose 1 point for Type II errors, lose K points for Type I errors.

$K = 19 \Rightarrow$ one-sided test with 5% level is minimax.
$K = 99 \Rightarrow$ one-sided test with 1% level is minimax.

Problems:

- Generates hypothesis testing rules, but not the criterion
- Big errors and tiny errors are treated the same
- Is 5% used because Type I errors are always 19 times worse?

Testing as a game against nature

- Nature picks θ (the treatment effect)
- Statistician observes a noisy estimate $\hat{\theta} \to \theta$.
- What if the statistician has no prior about the way nature picks θ?

Minimax criterion (aka maximin) $\Longrightarrow$ never approve innovations.
(Manski, 2004)

Minimax-regret criterion $\Longrightarrow$ accept if $\hat{\theta} > 0$ (50% test level)
Manski (2004), Hirano and Porter (2009), Schlag (2007), Stoye (2009)

Loss aversion with a factor of **102** under minimax-regret criterion could rationalize one-sided 5% level tests. (Tetenov, 2012)

Cannot be easily rationalized by typical nonlinear welfare functions.
(Manski and Tetenov, 2007)

Basic setup

One-shot game between a proponent and a regulator (no reputation).

Proponent has an idea for a new treatment/policy.
$\theta \in \Theta$ is the parameter capturing its quality,
known to the proponent, but not to the regulator.

$v(\theta)$ is the regulator's payoff if the proposal is approved. 0 if rejected.

$b(\theta) > 0$ is the proponent's payoff if approved. 0 if rejected.

Proponent could spend $\mathbf{c}$ to collect data $\mathbf{X} \in \mathcal{X}$ distributed $F(X;\theta)$.
- trial cost $\mathbf{c}$ is sunk before X is observed.
- "entry" decision based on expected payoffs.

Regulator approves/rejects based on the data X
- focus on statistical decision rules, not on more general contracts.
- decision rule depends on $b(\theta)$, c, $F(X;\theta)$ - all known to both parties.

Overview of the game with perfectly informed proponents

Timing of the game

- Regulator commits to a statistical decision rule δ according to which data will be mapped into acceptance decisions.
- Proponent learns his type $\theta \in \Theta$ (unknown to the regulator).
- Proponent chooses $\{$trial, no trial$\}$ whether spend c to collect evidence.
- Nature draws data X according to distribution $F(X;\theta)$ if trial. Both parties learn X.
- Regulator implements decision $\delta(X)$.

Payoffs to (proponent, regulator):

- $(0,0)$ if no trial
- $(-c,0)$ if trial and reject
- $(b(\theta)-c, v(\theta))$ if trial and approve

Common knowledge: trial cost c, payoffs $b(\theta)$, $v(\theta)$, distribution $F(X;\theta)$.

The regulator commits to a statistical decision rule:

$$\delta : \mathcal{X} \to [0, 1].$$

$\delta(X) = 0$: reject when the data is X, $\quad \delta(X) = 1$: accept.

Prior to the clinical trial, the probability that an innovation with value θ would be accepted is

$$\beta_\delta(\theta) \equiv \int_{\mathcal{X}} \delta(X) dF(X; \theta).$$

In statistics, $\beta_\delta(\theta)$ is the **power function** of test δ.

Acceptance probability drives the proponent's decision to collect data.

(Risk-neutral) proponent's best response to δ:

$$\beta_\delta(\theta) > \frac{c}{b(\theta)} \implies \text{conduct the trial,}$$
$$\beta_\delta(\theta) < \frac{c}{b(\theta)} \implies \text{no trial}$$

Because of commitment, we could study the regulator's single-agent decision problem, taking into account the proponent's best response.

The regulator's payoffs are

$$\begin{aligned} v(\theta) \cdot \beta_\delta(\theta) \quad &\text{if} \quad \beta_\delta(\theta) > \frac{c}{b(\theta)} \\ 0 \quad &\text{if} \quad \beta_\delta(\theta) < \frac{c}{b(\theta)} \end{aligned}$$

To attain maximum payoff for $v(\theta) < 0$, it is sufficient to set $\beta_\delta(\theta) < \frac{c}{b(\theta)}$.

If the decision to conduct a trial is "exogenous," the regulator has to set $\beta_\delta(\theta) = 0$ (no approvals) to achieve the same payoffs for $v(\theta) < 0$.

There's a substantial difference in the supply of ideas with $\theta < 0$ and $\theta > 0$:

> *"Discovery consists precisely in not constructing useless combinations, but in constructing those that are useful, which are an infinitely small minority."*
>
> Henri Poincare, *Science and Method*

Null hypothesis: Θ_0: $v(\theta) < 0$.
It's easy to propose treatments that are worse than the status quo. If there were positive expected profits for proposing and testing ideas with $v(\theta) < 0$, everyone could try.
Worst-case prior $P(\Theta_0) \to 1$ is quite reasonable.

Alternative hypothesis: $v(\theta) > 0$.
Beneficial innovations are in an "infinitely small minority."

Fully deterrent tests

Proposition 1 Decision rules δ^* that control test size:

$$\beta_{\delta^*}(\theta) < \frac{c}{b(\theta)} \quad \forall \theta \in \Theta_0$$

are minimax for the regulator w.r.t. θ.

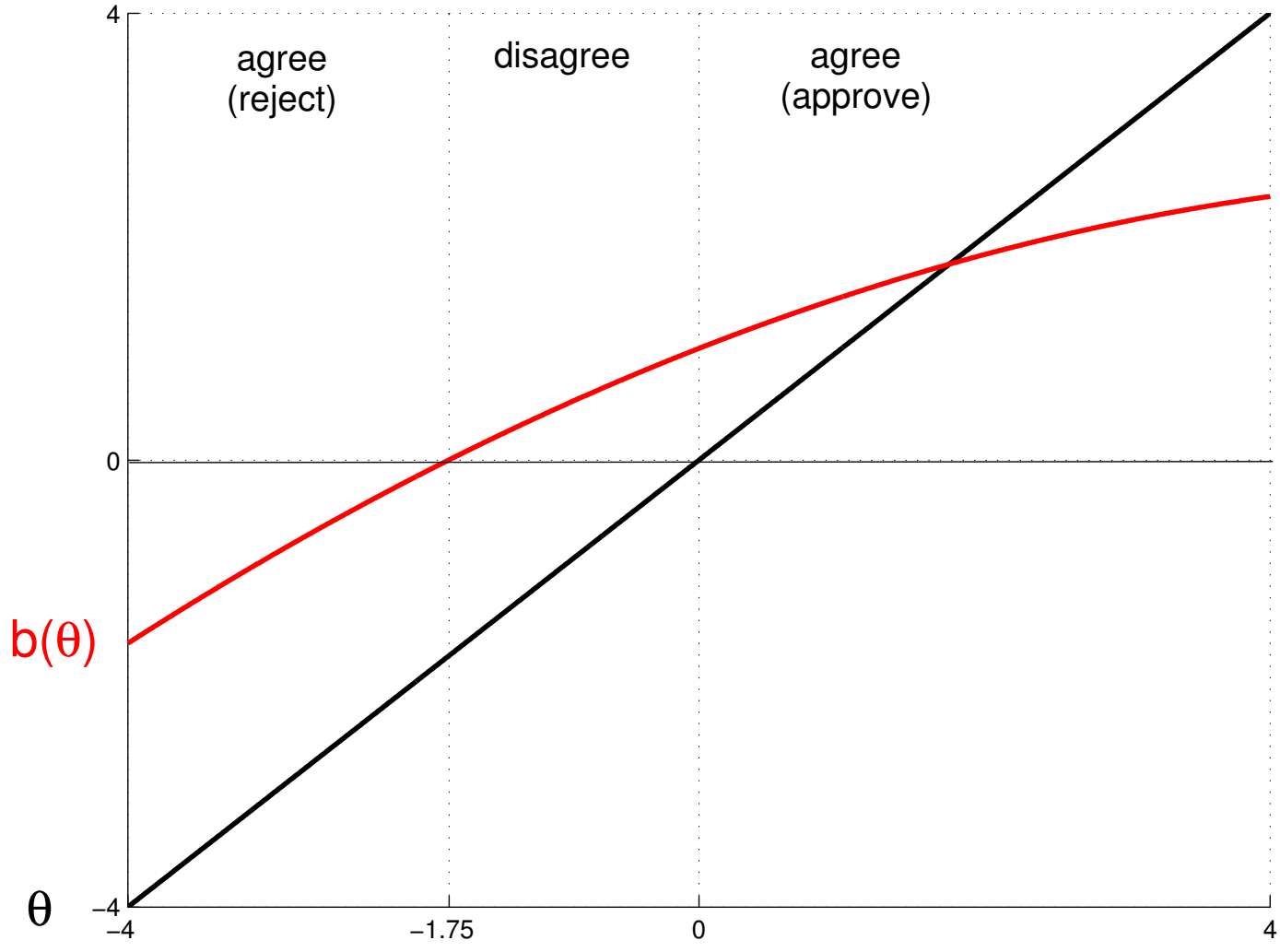
In the simple case of $b(\theta) = b$, this yields the classical hypothesis testing criterion with level $\frac{c}{b}$.

Among such decision rules, the regulator could try maximizing power (probability of acceptance) over $\Theta_1 : v(\theta) > 0$.

Proponents with precise information

Add structure to compare the fully deterrent test with optimal solutions of a Bayesian regulator who has a prior on θ

- $\theta \in \mathbb{R}$
- $v(\theta) = \theta$: θ is the net value of the proposal to the regulator.
- $F(X; \theta)$ is continuous and satisfies the Monotone Likelihood Ratio property.
  Leading example $X \sim \mathcal{N}(\theta, \sigma^2)$, known σ^2.
- Proponent's benefit is a continuous non-decreasing function $b(\theta) > 0$.


agree
(reject)
disagree
agree
(approve)
4
0
−4
b(θ)
θ
−4
−1.75
0
4

Proponents with precise information

The regulator could consider only monotone (threshold) decision rules:

$$\delta_T(X) = \begin{cases} 0 \text{ for } X < T, \\ 1 \text{ for } X \geq T. \end{cases}$$

because any decision rule could be replaced by a monotone one which preserves $\beta_\delta(0)$, doesn't reduce $\beta_\delta(\theta)$ for $\theta > 0$ and doesn't increase $\beta_\delta(\theta)$ for $\theta < 0$.
(Karlin and Rubin, 1956)

Monotone decision rules could be ordered by the threshold T and correspond to one-sided tests of different sizes.

There is a threshold decision rule δ^* for which

$$\beta_{\delta^*}(0) = \frac{c}{b(0)}$$

Will call it the **fully deterrent test**.

Then for all $\theta < 0$ it is not profitable to conduct trials

$$\beta_{\delta^*}(\theta) \cdot b(\theta) < \beta_{\delta^*}(0) \cdot b(0) = c$$

while for all $\theta > 0$ it is.

Proposition 2
δ^* is admissible (there's no decision rule at least as good for all θ and strictly better for some θ) and minimax.

δ^* is the only admissible minimax decision rule.

Higher threshold (lower test size) makes the rule inadmissible. It has a strictly lower acceptance probability (hence lower payoff to the regulator) for all $\theta > 0$. It has the same payoff for $\theta < 0$.

Lower threshold (higher test size) rules are not minimax, the regulator's payoff is negative for some $\theta < 0$, which is lower than the minimum payoff of δ^* (which is zero).

Multiple trials

Proponents have to pay the trial costs before observing the outcome.

If playing once isn't profitable for them, playing many times and picking the best result also isn't profitable.

Certain proponents with $\theta > 0$ who get a low value of X and do not get acceptance would find it profitable to retry (with the same $c, F, b(\cdot)$).

Comparison with Bayesian regulators

A testing rule that deters all proponents with $\theta < 0$ from trials is too strict for a Bayesian regulator.

Suppose the regulator has a prior distribution $Q(\theta)$ on potential proponent types.

Optimal tests are not from updating the prior $Q(\theta)$, i.e.,

$$\max_T \int \theta \beta_{\delta_T}(\theta) dQ(\theta)$$

Bayesian regulator's problem accounting for the self-selection of proponents is:

$$\max_T \int \theta \beta_{\delta_T}(\theta) \cdot \mathrm{I}\left[\beta_{\delta_T}(\theta) b(\theta) \geq c\right] dQ(\theta)$$

Proposition 3
A Bayesian regulator's decision rule will always set a lower evidence threshold than the fully deterrent test.

Hence, some range of proponents with slightly bad ideas $\bar{\theta} < \theta < 0$ will find it profitable to try them out (and some of them will be approved). In exchange, all good ideas have a higher probability of acceptance.

Proposition 4:
If you consider priors Q_n with $Q_n(\theta < 0) \to 1$ and positive density on $[-\epsilon, 0]$, Bayesian regulator's decision rules will converge to the fully deterrent test rule.

Bayes vs Minimax

Hypothesis testing with level $\frac{c}{b(0)}$ is close to optimal if the regulator is pessimistic about the distribution of potential proposals $Q(\theta)$.

Truncated part of the distribution of potential proposals is completely unobservable if some testing procedures are already in place, making it hard to have an "informed prior"

As good ideas are implemented, coming up with additional improvements may be harder.

Proponents uncertain about θ

Proponent has a prior distribution π on $\theta \in \mathbb{R}$.

Regulator doesn't know π and doesn't have a prior about it.

Regulator considers proponent's beliefs "rational" - the regulator would use π if these beliefs were revealed.

Common knowledge: cost of data $\mathbf{c}$ and proponent's gain from approval $\mathbf{b(0)}$.

Results in this case rely on additional assumptions:

- Proponent's payoff $b(\theta)$ is concave in θ.
- The ratio $\frac{-\frac{dF(T;\theta)}{d\theta}}{1-F(T;\theta)}$ is non-increasing in θ for all T.
  Examples: normally or exponentially distributed X.

Since θ could be negative, the results of the trial may convince the proponent not to seek regulatory approval. The ex ante probability that both parties agree on approval is $\beta_{\delta,\pi}(\theta)$.

Proposition 5

If the regulator's expected payoff (w.r.t. π) conditional on the proponent collecting evidence is negative

$$\int_{\mathbb{R}} \theta \beta_{\delta,\pi}(\theta) d\pi(\theta) < 0,$$

then it is not optimal for the proponent to conduct the trial:

$$\int_{\mathbb{R}} b(\theta) \beta_{\delta,\pi}(\theta) d\pi(\theta) - c < 0.$$

Proposition 6

Hypothesis test rule δ^* with fully deterrent test size

$$\beta_{\delta^*}(0) = \frac{c}{b(0)}$$

is admissible and minimax with respect to π.

Choice of trial costs and precision

The fully deterrent test rule could be applied for any trial design (c, F) chosen by the proponent (as long as (c, F) is known to the regulator).

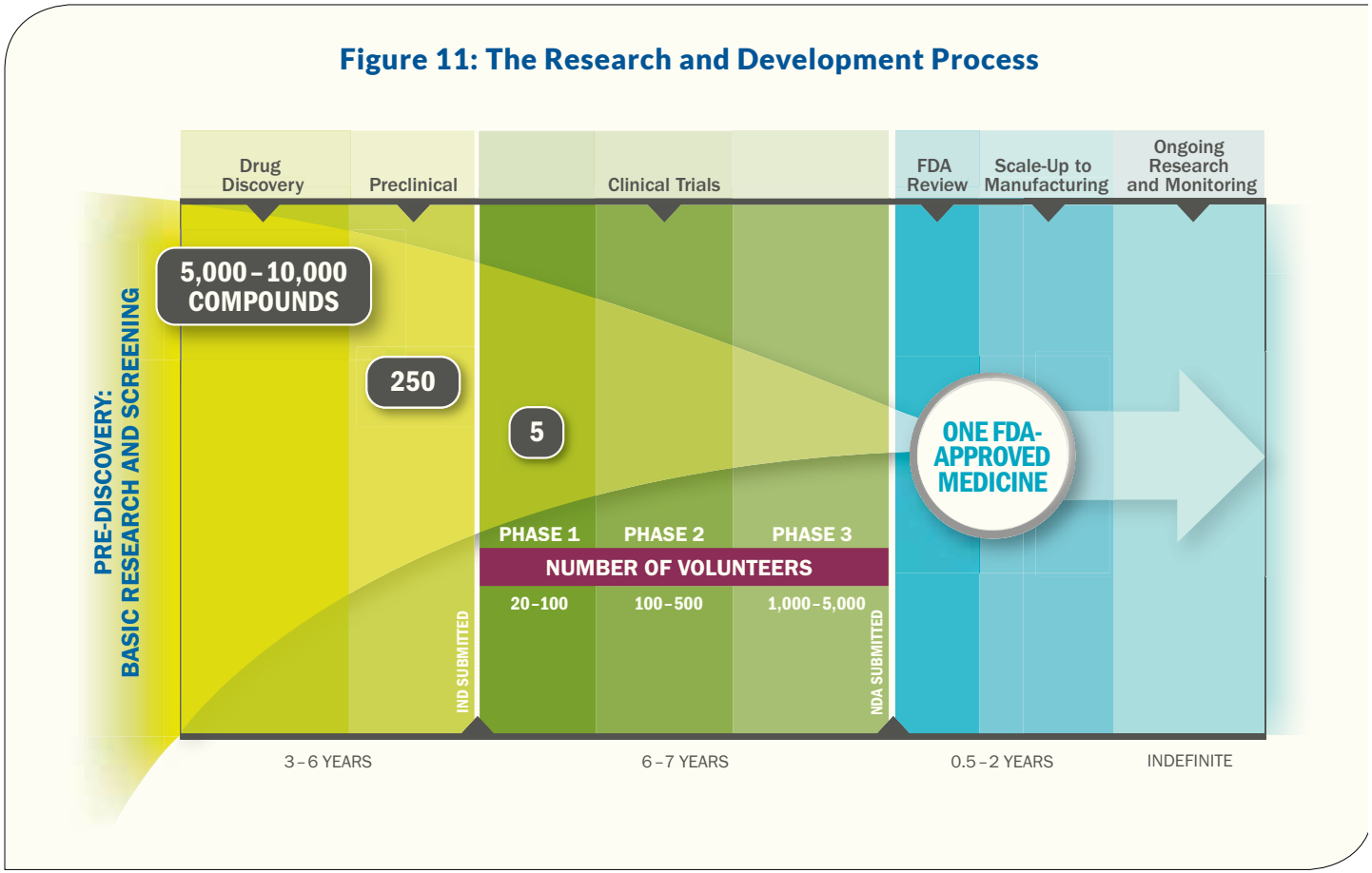
Choice of (c, F) creates complicated incentives for the regulator:

- Regulator may want to be stricter for some trial designs in order to induce a different choice of (c, F)
- Regulator may accept less precise experiments to make entry sufficiently profitable for some types of proponents.

Open question: is a hypothesis test rule with level $\frac{c}{b(0)}$ for any proponent's choice of (c, F) admissible or should some choices of trial design (c, F) always be discouraged?

Illustration: Phase III clinical trials overview

Last stage of clinical trials before drug approval.
Closest to an ideal randomized experiment.
Well documented.
Very expensive (36% of annual R&D expenses in 2011).



Reproduced from: 2013 Biopharmaceutical Research Industry Profile (PhRMA)

Phase III clinical trials: costs and benefits

Costs of Phase III clinical trials are spread over 2-3 years.

Sales are spread over 20+ years.

Both need to be discounted to the start of the trials
(could discount to any other date if we're interested in their ratio).

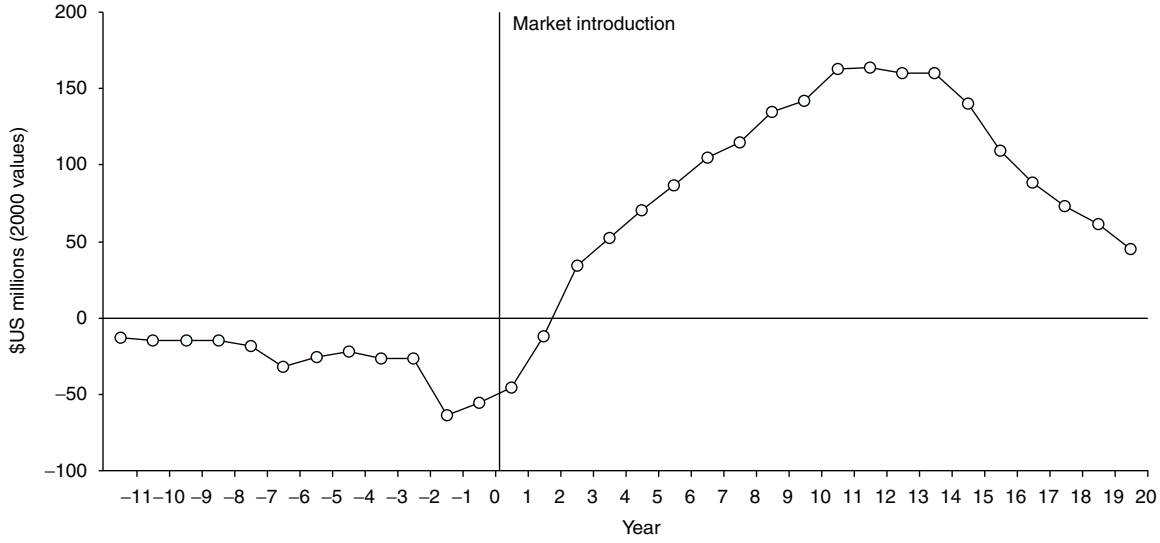


Fig. 5. Cash flows over the product life cycle: baseline case.

Reproduced from: Grabowski, Vernon and DiMasi (2002)

Phase III clinical trials: test level

Fully deterrent test level for drug i: $\alpha_i = \frac{c_i}{b_i(0)}$.

c_i = present value of expected Phase III clinical trial costs

$b_i(0)$ = present value of expected profits "unlocked" by the approval if $\theta_i = 0$.

Both vary a lot.

Don't have such data for individual drugs.

Phase III clinical trials: representative drug

Will consider a "representative" drug with:

c = average cost of conducted Phase III trials.
$b(0)$ = average profit of approved drugs.

Data source: DiMasi et al. (2003), summary data on R&D expenses by phase of development from a confidential survey of firms.

Fully deterrent test level for a representative drug:

$$\alpha = \frac{\$119.2 \text{ million}}{\$802 \text{ million}} = 14.9\%.$$

\$802 mln. = average P.V. of pre-approval R&D expenses per approved drug.

Grabowski et al. (2002) analyze sales data for the earlier half of DiMasi et al. sample and find that average R&D expenses $\approx$ average profits.

Phase III clinical trials: variability

Need to know joint distribution of $(c_i, b_i(0))$ to find out the distribution of deterrent test levels.

Drugs in the top decile have 5.5 times higher average sales.

Assuming average clinical trial expenses, test level for a top-decile drug:

$$\alpha == \frac{\$119.2 \text{ million}}{5.5 \cdot \$802 \text{ million}} = 2.7\%.$$

If approval depended only on a single test, conventional levels of 5% and 1% would be a strong deterrent.

Regulator tied to using conventional test levels could adjust c and $b(\cdot)$ instead. *Orphan Drug Act* tried to effectively reduce c and increased b for drugs targeting rare conditions.

Conclusion

Controlling the level of a hypothesis test may be an economically rational strategy for deterring null hypotheses from being tested.

The level of the test is dictated by the economic parameters, not by convention.

Self-interested response to statistical procedures could be an important consideration that could be used to design them.