

Intencionalidade vs. Randomização: Eficiência vs. Viés ou Tendenciosidade

Por que (não) Sortear / Aleatorizar?

Julio Michael Stern

IME-USP - Universidade de São Paulo



<http://www.ime.usp.br/~jstern/miscellanea/jmsslide/ABJ161.pdf>
– Superior Tribunal Eleitoral, Brasília, 7-9 de Agosto de 2016
– Hospital do Cancer A.C. Camargo, Outubro de 2016

0b- Esta apresentação

0- Introdução

- 1- Charles Sanders Peirce, Ronald Fisher e o Padrão Ouro: Ensaios (duplo) cegos e randomizados
- 2- Confundimento e Separação / Desacoplamento
- 3- Esquema Box & Hunter suas Maldições
- 4- Ensaios Clínicos e Legais: Considerações Éticas
- 5- Amostragem Intencional Fortuita (Haphazard), Protocolos dinâmicos para ensaios complexos, etc.
- 6- Niklas Luhmann: (De-) Diferenciação Sistemática.
Jurimetria: Colaboração Inter-disciplinar + Respeito a valores, especificidades e competências Intra-disciplinares

História dos contextos de desenvolvimento: Testes psicométricos; Ensaios clínicos, Experimentos de política pública.

1a- Testes Psicométricos

Entre 1850 e 1880 a psicofísica (psychophysics) tentava quantificar a resposta humana a estímulos sensoriais. Por exemplo, a Lei de Fechner afirma que:

– *A magnitude da sensação (γ) percebida é proporcional ao logaritmo da magnitude do estímulo (β) expresso em termos do limiar de percepção (b), abaixo do qual este se torna imperceptível, isto é, $\gamma = k \log(\beta/b)$ ($\beta > b$).*

Para testar a validade se sua (própria) teoria, Gustav Fechner fazia experimentos nos qual ele (mesmo) preparava dois estímulos, de intensidade μ e $\mu + \delta$. Posteriormente, ele (ele mesmo, novamente) se submetia a ambos os estímulos para verificar se era possível (ou não) distinguir qual o mais forte.

Introspecção: Meditação para evitar conflitos de interesse.

1b- Testes Psicométricos

Em 1870, Georg Müller e sua equipe em Tübingen e Göttingen estabelecem a necessidade de fazer Ensaio Cegos:

- Os estímulos são preparados e administrados por um agente, o Operador ou Experimentador, e aplicados em outro agente, o Paciente ou Sujeito, que Não conhece a forma de preparo.
- Os estímulos são apresentados em “ordem irregular” (buntem Wechsel). Esta irregularidade visa impedir que o paciente se habitue ou seja capaz de perceber padrões de regularidade na seqüência de estímulos apresentados ou, em outras palavras, evitar que o paciente forme expectativas que lhe ajudem a adivinhar a resposta correta.

Todavia, não havia ainda nem uma teoria definindo Irregularidade nem um método sistemático para obter seqüências irregulares.

1c- Ensaios Randomizados

Em 1885, Charles Saunders Peirce e seu estudante Joseph Jastrow, no artigo *On small Differences of Sensation*, (Mem. Nat. Acad.of Sciences, 3, (1884), p.75-83), apresenta a Randomização como solução prática para obter uma seqüência irregular, utilizando com instrumento de aleatorização um baralho normal de cartas de jogo.

A ideia de Peirce foi (na época) violentamente rejeitada! Uma método científico não poderia valer-se de protocolos envolvendo sorteios aleatórios e utilizando cartas de jogo!

Em 1926, Ronald Fisher resgata o conceito de experimento randomizado. A partir de então, o ensaio cego e randomizado passa a ser o padrão ouro da ciência experimental.

1d- Boas Seqüências Aleatórias

*Em computadores eletrônicos digitais, é muito conveniente calcular uma seqüência aleatória um número por vez, no tempo em que requerido, através de uma regra bem especificada (algoritmo determinístico) que, no entanto, é concebida de forma que nenhum teste estatístico **razoável** seja capaz de detectar um desvio significativo da randomicidade. Uma seqüência assim concebida é denominada pseudo-aleatória. Hammersley and Handscomb (1964, p.26)*

*Uma seqüência de números pseudo-aleatórios, U_i , é uma seqüência determinística em $[0, 1]$ tendo as mesmas propriedades estatísticas **relevantes** que uma seqüência de números aleatórios. Ripley (1987, p.15).*

2a- Confundimento e Separação

Tablela1: Paradoxo de Simpson.

Sexo	T	R	NR	Tot	R%
Todos	T	20	20	40	50%
Todos	NT	16	24	40	40%
Masc.	T	18	12	30	60%
Masc.	NT	7	3	10	70%
Fem.	T	2	8	10	20%
Fem.	NT	9	21	30	30%

Lindley (1991, p.47-48): De 80 pacientes, 40 são tratados (T), 40 não (NT ou placebo) . Alguns pacientes se recuperam (R), outros não (NR). A taxa de recuperação (R%) é dada na tabela para todos os pacientes de forma agregada e desagregada por sexo. Aparentemente, o tratamento é ruim tanto para homens quanto para mulheres, mas é bom para todos!

2b- Causa e Efeito do Confundimento

Lindley (1991, p.47-48):

O médico responsável pelo ensaio:

- Não confiava no tratamento, e
- Sabia que a doença afeta mulheres mais severamente.

Assim, ele decidiu testar o tratamento (T) apenas nos homens, que provavelmente se recuperariam de qualquer forma.

- Na linguagem de ensaios clínicos, não tratamento (NT) ou placebo denota um procedimento inócuo ou a droga padrão, enquanto tratamento (T) denota a novidade proposta.

Assim procedendo, o médico criou uma **associação expúria** entre a variável explicada (taxa de recuperação) e uma variável explicativa (sexo).

2c- Separação / Desacoplamento

Um experimento randomizado de Fisher... consiste de duas partes: Randomização e Intervenção.

Intervenção significa que mudamos o comportamento natural do indivíduo: Separamos sujeitos em dois grupos, chamados Tratamento e Controle, e convencemos os sujeitos a obedecer o protocolo do experimento. Damos tratamento a alguns pacientes que, em circunstâncias normais, não o buscariam, e damos placebo a pacientes que de outra forma buscariam tratamento. A isto, em nosso vocabulário, denominamos 'cirurgia' - estamos cortando um vínculo funcional e substituindo-o por outro. A grande intuição (insight) de Fisher foi que conectar o novo vínculo a uma moeda aleatória 'garante' que o vínculo que queríamos quebrar foi de fato quebrado. A razão é que assumimos que a moeda não é afetada por nada que possa ser medido a nível macroscópico. Pearl (2000, p.348, The Art and Science of Cause and Effect)

2d- Confundimento Complexo (efeitos de 2a ordem)

Caso Gluconato de Zinco

– Placebo casado (placebo matching): Um estudo cego requer um placebo com as mesmas características (cor, gosto, cheiro, consistência, apresentação) que a droga em teste.

– *Qualquer coisa de gosto tão ruim e que fica na boca tanto tempo como o gosto de zinco, tem que ser um bom remédio!*
Desbiens (2000), Farr and Gwaltney (1987).

– Pacientes ou participantes de um ensaio clínico não deveriam interagir fora do protocolo estabelecido. No entanto eles o fazem, e depois escondem e mentem sobre o que fizeram!

– Semelhante a regras de interação entre membros do júri.

– As regras e protocolos de um bom experimento devem, necessariamente, levar em conta efeitos de ordem superior, ainda que (ou principalmente se) indesejados.

3a- Esquema Box & Hunter suas Maldições

Box & Hunter (1978). Statistics for Experimenters:
– Controle o que puder e randomize o resto!

Maldição da dimensionalidade

Falta / raridade de irmãos gêmeos

Maldição assintótica

4a- Conflitos Éticos com a Ciência

Conflito de interesse no caso do Paradoxo de Simpson

- Ética médica: Tratar o paciente da melhor forma possível.
 - Ética científica: Obter a melhor informação possível.
- ⇒* Necessidade de ensaio duplo-cego e randomizado.

Conflito de interesse no caso do Gluconato de Zinco

– Direito à Informação:

O paciente quer conhecer seu próprio tratamento.

– Veracidade, evitar viés e tendenciosidade:

É preciso esconder do paciente esta informação

⇒* Necessidade de ensaio multi-cego e randomizado.

* Eu sou um cientista, não sou médico, nem jurista.

* Ética médica: Fazer (apenas) aquilo que beneficia o paciente.

Ensaio clínico não pertencem a este escopo. No entanto, são absolutamente necessários para expandir o conhecimento da ciência e o escopo da prática médica.

4b- Conflitos Éticos externos à Ciência

- Caso Talidomida (thalidomide), com efeitos teratogênicos
- Caso Fosfoetanolamina, com fomento de crenças infundadas
⇒** Opção por fazer ensaios estatísticos seguros, isto é, amplos, abrangentes e altamente confiáveis, ainda que sejam grandes (muitos pacientes) e longos.
- Caso AIDS / HIV, com urgência de anti-retro-virais
⇒** Opção por fazer ensaios estatísticos rápidos, isto é, pequenos e curtos, assumindo os riscos conseqüentes.

** Dentro das normas de boa prática estatística, é possível desenhar ensaios para atender ambos os tipos de demanda. Obviamente, haverá sempre um conflito de escolha e conseqüente compromisso (trade-off) entre qualidade e custo da informação.

4c- Efeitos (não tão) éticos de 2a ordem

- » Agencias reguladoras (NIH, ANVISA) são conservadoras e avessas ao risco, dada sua grande responsabilidade; (Princípio de legalidade estrita?!)
- » Induzir altos custos de desenvolvimento;
- » Desenvolver apenas projetos suficientemente lucrativos;
- » Empurrar grandes ensaios para jurisdições de menor custo operacional e baixo risco jurídico;
- » ...

5- Soluções* Tecnológicas p/ Conflitos Éticos

- Amostragem Intencional Fortuita
 - Protocolos Dinâmicos
-

* Obviamente, são paliativos, não soluções, vide slide 6b

Christoph Engel
Wim Voermans

6b- (De-) Diferenciação Sistêmica

Luhmann

- J.M. Stern (2008). Cognitive Constructivism and the Epistemic Significance of Sharp Statistical Hypotheses in Natural Sciences.
- N.A. Desbiens (2000). Lessons Learned from Attempts To Establish the Blind in Placebo-Controlled Trials of Zinc for the Common Cold. *Ann Intern Med.* 133, 302-303.
- Ch. Engel, H. Hennig-Schmidt, B. Irlenbusch, S. Kube (2009). On Probation: An Experimental Analysis. Max Planck Institute for Research on Collective Goods at Bonn, Tech.rep. 2009/38.
- Ch. Engel (2013). Behavioral Law and Economics: Empirical Methods. Max Planck Institute for Research on Collective Goods at Bonn, Tech.rep. 2013/1.
- Ch. Engel (2013). Legal Experiments – Mission Impossible? Inaugural lecture of the Sanders Chair for Internationalisation at Erasmus University School of Law at Rotterdam.
- Ch. Engel (2015). Randomized Information about the Law as an Instrument. *JITE – Journal of Institutional and Theoretical Economics*, 171, 1, 171-175.
- B.F. Farr, J.M. Gwaltney (1987). The problems of taste in placebo matching: an evaluation of zinc gluconate for the common cold. *Journal of chronic diseases* 40, 9, 875-879.
- T.J.Kaptchuk, C.E.Kerr (2004). Commentary: Unbiased Divination, Unbiased Evidence, and the Patulin Clinical Trial. *International Journal of Epidemiology*, 33, 247-251.
- Kotz et al. (2005), under the entries Clinical Trials I, by N.E.Breslow, v.2, p.981-989, and Clinical Trials II, by R.Simon, v.2, p.989-998.
- S.Kotz, N.Balakrishnan, C.B.Read, B.Vidakovic, eds. (2005). *Encyclopedia of Statistical Sciences*, 2nd ed. Wiley-Interscience. Entries: Clinical Trials I, by N.E.Breslow, v.2, p.981-989, and Clinical Trials II, by R.Simon, v.2, p.989-998.
- N. Luhmann (1989). *Ecological Communication*. Chicago Univ. Press.
- N. Luhmann (1995). *Social Systems*. Stanford Univ. Press.
- J.H.Noseworthy, G.C.Ebers, M.K. Vandervoort, R.E.Farquhar, E.Yetisir, R.Roberts (1994). The impact of blinding on the results of a randomized, placebo-controlled multiple sclerosis clinical trial. *Neurology*, 44, 16-20.
- W. Voermans (2014). Motive-Based Enforcement. Proc. of the 10th Cong. of the International Association of Legislation (IAL). Baden-Baden: Nomos p.41-61.
- W. Voermans (2015). To Measure is to Know: The Quantification of Regulation. *The Theory and Practice of Legislation*, 3, 1, 91-111.
- M. Wan, M. Orlu-Gul, H. Legay, C. Tuleu (2013). Blinding in pharmacological trials: the devil is in the details. *Archives of disease in childhood*, archdischild-2013.