Chairperson(s): Jeong Min Lee *Seoul National University Hospital, Korea*

통계학전문가의 관점

Kijun Song
Yonsei University College of Medicine, Korea. biostat@yuhs.ac

Randomized Clinical Trial: Why? - Statistical perspectives

송 기 준
연세대학교 의과대학 의학통계학과

Evidence-Based Medicine (EBM)

- Ideally based on data from clinical research
- Must understand fundamentals of design and analysis
- Randomized controlled clinical trial (RCT) is gold standard.

연구 목적에 따른 임상시험

확증 시험(Confirmatory Trial)	탐색 시험(Exploratory Trial)
▪ 가설을 미리 제시하고 평가하는 통제된 임상시험 ▪ 어떤 주장을 지지하기 위한 확실한 증거 제시 ▪ 임상시험계획서와 업무지침서(SOP) 준수 필수 ▪ 임상시험 설계에 대한 정당성 및 분석원칙 명시	▪ 미리 정의된 가설을 검정하는 것이 주 목적이 아님. ▪ 자료 탐색 ▪ 시험을 통해 새로운 가설 형성 ▪ 유효성을 공식적으로 입증하지 못함.

[임상시험통계지침서 中]

Bias를 줄이기 위한 방법

- 눈가림(맹검; blinding or masking)
  - 임상시험을 시행하거나 자료를 분석할 때 피험자가 어느 군에 배정되었는지 알지 못하게 하는 것
  - 눈가림의 유형
    - No blinding(open trial), Single blinding, Double blinding, Triple blinding
- 무작위배정(randomization)
  - 피험자를 각 치료군에 무작위로(확률적으로) 배정하는 것
  - 눈가림과 함께 무작위배정을 사용하면 치료군 배정을 예측함으로써 발생할 수 있는 피험자 선정 및 배정에 관한 bias를 제거할 수 있음.

Advantages of Randomization (1)

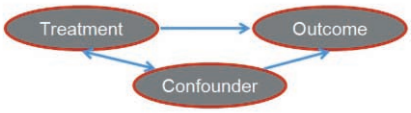
- Randomization "tends" to produce comparable groups.

Design	Bias
Randomized control	Chance
(Non-randomized) Concurrent control	Chance & selection bias
(Non-randomized) Historical control	Chance , selection , & time bias

→ Thus, randomization makes the comparison between the test and control groups more credible.

Advantages of Randomization (2)

- Randomization can make the confounder independent of the treatment assignment.

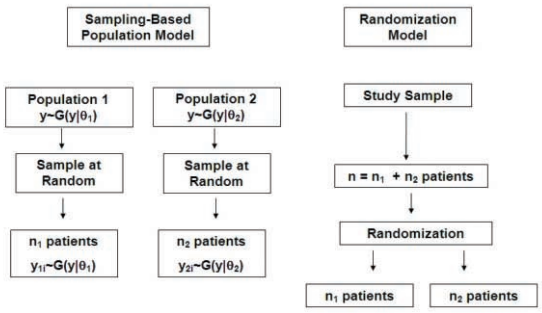


– In non-randomized study, potential confounding factors are usually adjusted by using statistical models(any regression models).
→ However, this estimate of treatment effect is incorrect or imprecise.
– Randomization can eliminate the effect of confounder.
→ So, any observed association between treatment and outcome gets nearer to causality.

Advantages of Randomization (3)

- Randomization allows for valid statistical tests.
  - Randomization justifies the analysis typically conducted without depending on external distribution assumptions.
    - → Randomization procedure is associated with the "randomization test" that does not rely on the assumptions of distribution.
    - → Common tests such as t-tests and chi-square tests are approximations to the randomization test.
  - Comparison of non-randomized groups is based on "population model", but that of randomized groups is based on "randomization model".

Population model vs. Randomization model



Randomization test(Permutation test): Example

(Journal of Magnetic Resonance Imaging 2009;30:621–630)

MR Imaging of Endometrial Carcinoma for Preoperative Staging at 3.0 T: Comparison With Imaging at 1.5 T

Masatoshi Hori, MD, PhD,[1]* Tonsok Kim, MD, PhD,[1] Takamichi Murakami, MD, PhD,[2] Izumi Imaoka, MD, PhD,[2] Hiromitsu Onishi, MD, PhD,[1] Atsushi Nakamoto, MD,[1] Yasuhiro Nakaya, MD, PhD,[1] Kaname Tomoda, MD, PhD,[1] Tateki Tsutsui, MD, PhD,[3] Takayuki Enomoto, MD, PhD,[3] Tadashi Kimura, MD, PhD[3] and Hironobu Nakamura, MD, PhD[1]

Purpose: To prospectively compare magnetic resonance imaging (MRI) at 3.0 T and 1.5 T in the same patients for preoperative evaluation of endometrial carcinoma.

Materials and Methods: Thirty consecutive patients with endometrial carcinoma underwent MRI at both 3.0 T and 1.5 T as well as surgery. Quantitative and qualitative analyses were performed. Two radiologists independently evaluated images. MR findings were compared with surgicopathologic findings.

Results: Image homogeneity of T2-weighted images at 3.0 T was significantly inferior to that at 1.5 T ($P = 0.007$). The scores of image homogeneity and susceptibility artifacts were not significantly different between 3.0 T gadolinium-enhanced imaging and 1.5 T imaging ($P = 0.09$ and 0.36). Kappa statistics showed good interobserver agreement between the two radiologists for local-regional staging on T2-weighted images (κ: >0.6). The area under the receiver operating characteristic curve (Az) values for T2-weighted imaging in terms of myometrial invasion, cervical invasion, and lymph node metastases were 0.88 (3.0 T) versus 0.91 (1.5 T), 0.84 versus 0.83, and 0.94 versus 0.95 for reader 1, respectively. There were no significant differences between imaging at 3.0 T and at 1.5 T in Az values for either reader ($P > 0.35$).

Conclusion: 3.0 T MRI is an equivalent imaging modality to 1.5 T imaging for presurgical evaluation of endometrial carcinoma, although not significantly superior to 1.5 T imaging.

Key Words: magnetic resonance; high-field strength imaging; ST; pelvis; uterine neoplasm; diagnostic accuracy
J. Magn. Reson. Imaging 2009;30:621–630.

ENDOMETRIAL CARCINOMA is the most common gynecologic malignancy and the fourth most frequent cancer among females in the United States, with an estimated 40,100 new cases and an estimated 7,470 deaths in 2008 (1). Important prognostic factors are depth of myometrial invasion, histological type, histological grade, vascular space invasion, nodal involvement, and peritoneal cytology (2–4). Preoperative accurate evaluation of the tumor extent could greatly optimize surgical procedure and therapeutic strategy (5–7). The usefulness of magnetic resonance imaging (MRI) in the preoperative assessment of endometrial carcinoma is increasingly being recognized (8–11).

Recently, whole-body 3.0 T MR systems are being increasingly used in clinical settings. Theoretically, 3.0 T imaging offers higher signal-to-noise ratio (SNR) and enhanced spectral separation than does 1.5 T imaging. The gain in SNR can be maintained or traded for either speed or spatial resolution, or both. Compared with 1.5 T, potential outcomes of 3.0 T imaging include changes in tissue T1 and T2 relaxation times, increased mag-

Randomization test(Permutation test): Example

(Journal of Magnetic Resonance Imaging 2009;30:621–630)

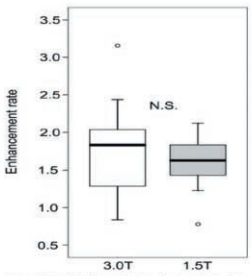


Figure 4. Boxplots show results for enhancement rate of the myometrium on contrast-enhanced T1-weighted images. The plots indicate the median (central horizontal line), the 75th quartile (top of box), the 25th quartile (bottom of box), and the smallest and largest nonextreme values (whiskers). There was no significant difference in mean values between 3.0 T MRI and 1.5 T MRI ($P = 0.45$, unpaired t-test). ○ = outliers.

Question ?

3.0T군과 1.5T군의 표본 평균의 차이가 원래 모집단의 평균에 차이가 있기 때문에 생긴 것인가, 아니면 두 모집단의 평균은 원래 같았지만 우연히 표본을 추출하는 과정에서 발생한 차이인가?

Permutation test

[예] Autosomal dominant polycystic kidney disease (ADPKD)환자와 정상인 각 20명을 대상으로 seminal vesicles 크기를 3D CT의 axial image로 측정한 자료

ADPKD 환자 (20명)	정상인 (20명)
1.40	1.95
1.75	1.40
1.55	1.50
2.00	2.55
…	…
1.25	1.30
1.10	1.85
1.55	2.35
1.70	1.95
1.95	2.20

- 귀무가설 (H_0)
  환자군과 정상군간에 seminal vesicles크기 차이 없음.
- 대립가설 (H_1)
  환자군과 정상군간에 seminal vesicles크기 차이 있음.

두 집단의 vesicles 평균 차이
= 정상군 평균 - 환자군 평균
= 1.67 − 1.51 = 0.16

Permutation test

- 두 모집단 간에 원래 차이가 없다면 ?

➡ 환자군과 정상군의 seminal vesicles 수치를 무작위로 섞어서 비교해도 차이가 없을 것이라고 예상할 수 있음.

- 100개의 표본을 무작위로 생성.

Group	Vesicles 1	Vesicles 2	Vesicles 3	Vesicles 4	Vesicles 5	…	Vesicles 96	Vesicles 97	Vesicles 98	Vesicles 99	Vesicles 100
환자군	1.25	1.23	1.58	1.45	1.59		1.65	1.64	1.19	1.67	1.69
환자군	1.19	2.02	1.49	2.06	1.86		1.23	1.78	1.91	1.88	1.25
환자군	1.03	1.53	1.69	1.35	1.51		1.77	1.39	1.40	1.06	1.72
환자군	1.33	1.07	1.81	1.00	1.10		1.87	1.55	1.19	2.14	1.55
환자군	2.02	1.03	1.47	1.70	1.62		1.72	1.47	1.74	1.07	1.38
…	…	…	…	…	…		…	…	…	…	…
환자군	2.27	1.26	1.75	1.77	1.67		1.84	1.54	1.46	1.39	1.69
환자군	0.93	1.65	1.64	1.65	1.85		1.79	1.53	1.09	1.54	1.29
환자군	1.20	1.39	1.75	1.35	1.90		1.40	1.29	1.18	1.85	1.41
환자군	1.73	1.43	1.30	1.35	2.01	…	1.65	1.32	1.18	1.37	1.22
정상군	1.93	1.84	1.27	1.52	1.47		1.16	1.68	0.57	1.38	2.28
정상군	1.46	2.12	1.16	2.21	1.93		1.33	1.43	1.31	1.22	1.60
정상군	1.06	1.81	1.16	1.28	1.19		1.32	2.18	1.82	1.95	1.55
정상군	1.89	2.14	0.88	1.75	1.79		0.85	1.55	0.61	1.56	0.80
정상군	1.92	1.78	0.30	1.69	1.65		1.18	2.71	0.95	1.91	1.95
…	…	…	…	…	…		…	…	…	…	…
정상군	1.89	1.63	1.66	2.41	1.60		0.52	1.10	1.43	1.69	1.11
정상군	2.10	2.37	1.39	1.98	2.20		1.29	2.08	2.49	1.81	2.00
정상군	1.28	1.05	2.18	1.87	0.84		0.71	1.82	1.12	2.10	1.43
정상군	1.31	1.74	1.40	1.11	1.60		1.95	1.74	2.17	1.22	1.44
정상군	1.46	1.71	1.52	1.59	1.43		0.82	2.71	1.46	2.10	2.55
두 집단 간 평균의 차이	0.16	0.11	0.15	0.17	0.10		0.16	0.17	0.01	0.03	0.22

Permutation test

① Permutation test

두 집단 간 차이의 절대값이 0.16보다 큰 경우는 전체 100개의 표본 중에서 16개가 나옴. 즉, p-value는 0.16이라고 할 수 있음.

② Independent t-test

Independent t-test결과, p-value는 0.17 이었음.

Types of Randomization

- Simple(complete) randomization
- Restricted randomization
  - Permuted block randomization
  - Biased coin randomization(Efron, 1971)
  - Urn randomization (Wei & Lachin, 1988)
  - Etc.
- Randomization for balancing on known covariates
  - Stratified randomization
  - Minimization method(Pocock & Simon)
  - Etc.
- Adaptive randomization
  - Response-adaptive randomizations
  - Design-driven response-adaptive randomizations
  - Etc.

References

- Kempthorne, O. Why randomize? Journal of Statistical Planning and Inference, 1977; 1: 1-25.
- Lachin, J. Statistical properties of randomization in clinical trials. Controlled Clinical Trials, 1988;9:289-311.
- Lachin, J. Randomization in clinical trials: Conclusions and recommendations. Controlled Clinical Trials, 1988;9:365-374.
- Cook, T. and DeMets, D. Introduction to Statistical Methods for Clinical Trials. Chapman & Hall/CRC, New York. 2008.
- Piantadosi, S. Clinical Trials: A Methodologic Perspective. Wiley and Sons, New York. 2005.