



תכנון ניסויים

וביתוח שונות

פרופ' קמיל פוקס

סמסטר ב', תשס"ז

תכנון ניסויים

ניסוי הוא מחקר בו ניתן לשלוט בתנאי הסביבה.
המטרה של הניסוי הוא לבחון איך התנאים
הנשלטים משפיעים על מדדים או תוצרים
חשובים.

מחקר תצפיתי, לעומת זאת, הוא מחקר בו
התנאים נוצרים באופן טבעי, מבלי שהחוקר יכול
לקבוע אותם.

ניתוח שונות

ניתוח שונות הוא אחד הכלים השימושיים ביותר
בניתוח סטטיסטי של נתונים.

ניתוח שונות הוא כלי ניתוח עיקרי במחקרים
המשווים מדדים רציפים על פני מספר קבוצות.
הוא מרחיב את מבחן t להשוואת 2 קבוצות
למחקרים עם מספר רב של קבוצות, עם קבוצות
שנוצרות ממספר גורמים שונים, ועם מבנה
מורכב של חלוקה לקבוצות.

ניתוח שונות שימושי בניסויים וגם במחקרים
תצפיתיים.

דוגמה 1: ניסוי קליני לטיפול בלאוקמיה

מטרת הניסוי לשפר את ההישרדות של ילדים הלוקים בלאוקמיה לימפטית חריפה.

הרופאים רוצים להשוות בין מספר משטרי טיפול שונים, ולא יודעים איזה משטר עדיף.

הרופאים יקבעו לכל ילד את הטיפול שיקבל וירצו בסוף הניסוי להשוות בין הטיפולים.

מספר סוגיות בתכנון הניסוי

איך לקבוע מה יהיה הטיפול של כל ילד.

איך לדאוג שקבוצות הטיפול יהיו "ברי השוואה הוגנת".

איך כדאי למדוד את "הצלחת" הטיפולים.

מהו גודל המדגם הדרוש לניסוי.

דוגמה 2: מחקר תצפיתי בלאוקמיה

מטרת המחקר להעריך את החשיבות של ספירת דם לבנה בעת אבחון המחלה כגורם פרוגנוסטי אצל ילדים הלוקים בלאוקמיה.

מודדים לכל ילד את ספירת הדם הלבנה בעת האבחון. כמובן שהספירה לא בשליטת החוקרים. כמו-כן, יתכן שילדים השונים זה מזה בספירת דם לבנה שונים **גם במשתנים חשובים נוספים**.

סוגיות כמו אפיון טיפול מוצלח ובחירת גודל מדגם נשאות חשובות.

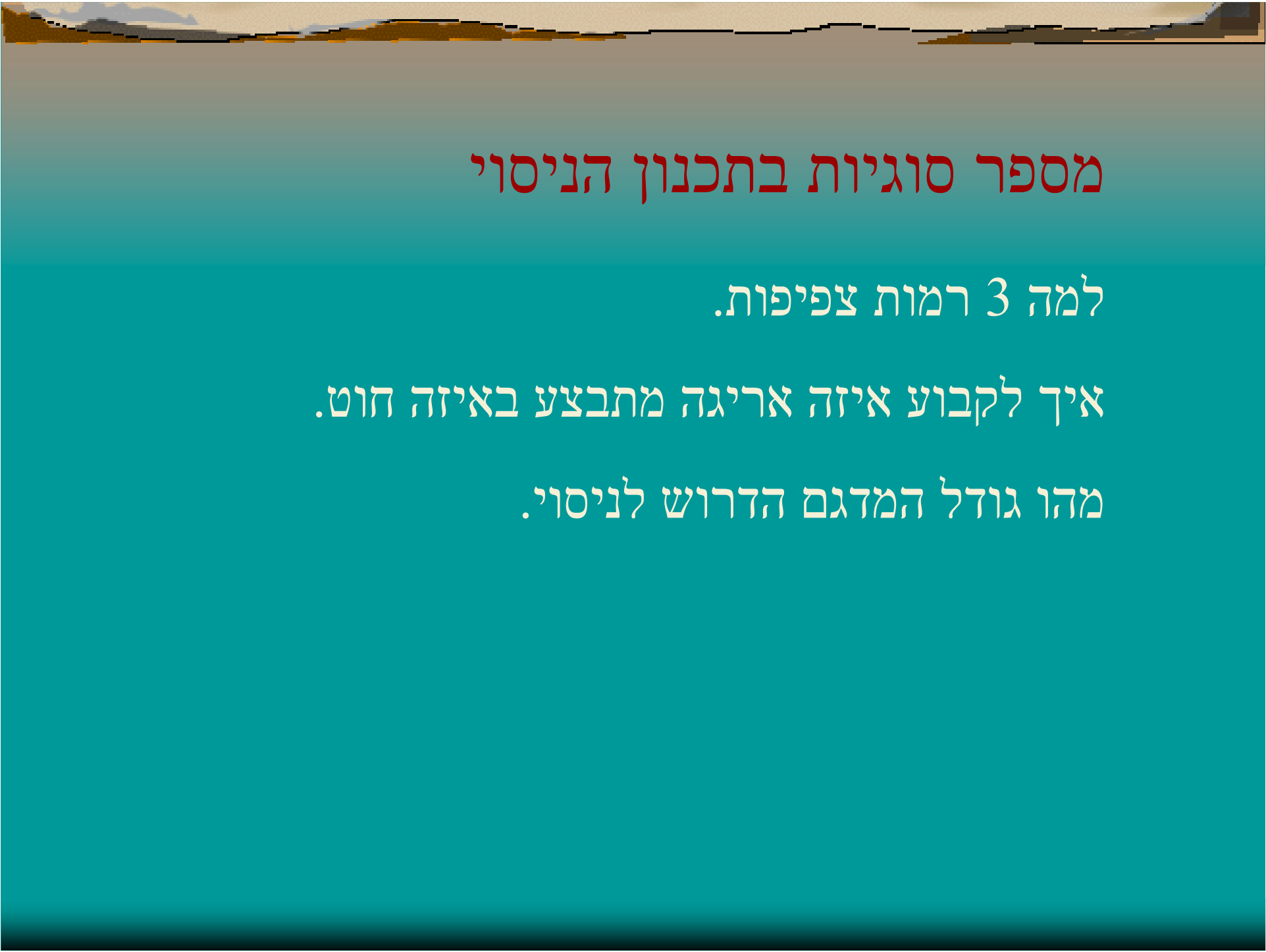
דוגמה 3: השוואת חוטים

מפעל לטקסטיל ערך ניסוי להשוות בין 6 סוגים של חוט.

החוטים הוכנו מ- 2 סוגים של כותנה וכל סוג הוכן ב- 3 רמות צפיפות: נמוכה, בינונית וגבוהה.

כל סוג של חוט נבדק ב- 9 אריגות.

המדד לטיב החוט היה מספר הפעמים שהחוט נקרע במהלך האריגה.



מספר סוגיות בתכנון הניסוי

למה 3 רמות צפיפות.

איך לקבוע איזה אריגה מתבצע באיזה חוט.

מהו גודל המדגם הדרוש לניסוי.

מרכיבים אופייניים לניסויים

- 1. יחידות ניסוי.** יחידות מאוד הומוגניות יכולות להדגיש השפעות. יחידות מגוונות יכולות לשפר את היכולת להכללה מן הניסוי.
- 2. טיפולים.** התנאים אותם נרצה לחקור. בניסויים רבים, "טיפול" אחד יהיה קבוצת ביקורת.
- 3. תוכנית ניסוי.** תוכנית שמפרטת איזה טיפול ניתן לכל יחידת ניסוי. חשוב לדאוג שלא יהיו הבדלים חשובים בין היחידות בכל קבוצת טיפול.

מרכיבים אופייניים לניסויים

4. בלוקים. לפעמים ניתן לחלק את יחידות הניסוי לקבוצות הומוגניות הנקראות **בלוקים (Blocks)**. אם כן, ניתן לנצל חלוקה זו בבניית תוכנית הניסוי. למשל, ספירת דם לבנה בעלת חשיבות פרוגנוסטית בליוקמיה. לכן, אפשר לחלק את הילדים לקבוצות על סמך ספירת הדם הלבנה שלהם ואז להקצות ילדים לטיפולים, בנפרד, מכל קבוצת ספירה.

מרכיבים אופייניים לניסויים

5. הקצאה אקראית. (Randomization) שיטה

מקובלת לקבוע את תוכנית הניסוי. הקצאה אקראית מונעת מן החוקר להטות את תוכנית הניסוי לטובת טיפול מסוים. היא גם מבטיחה שמשתנים נוספים יתחלקו באופן די שווה בין קבוצות הטיפול.

מרכיבים אופייניים לניסויים

6. **סמיות (Blind).** סמיות גם מצמצמת אפשרויות להטיה על

ידי כך שיחידות הניסוי וחוקרים לא יידעו מה היא קבוצת הטיפול של היחידה.

- **סמיות בודדה: (Single Blind)** יחידת הניסוי לא יודעת מה הטיפול.
- **סמיות כפולה: (Double Blind)** יחידת הניסוי לא יודעת מה הטיפול וגם החוקר המבצע את הטיפול לא יודע מה הטיפול.
- **סמיות משולשת: (Triple Blind)** יחידת הניסוי לא יודעת מה הטיפול וגם החוקר המבצע את הטיפול לא יודע מה הטיפול וגם החוקר המודד את התוצרים לא יודע מה הטיפול.

השוואה בין ניסוי מתוכנן ומחקר תצפיתי

ניסוי מתוכנן

1. התערבות יזומה
2. שליטה בתנאים
3. צורך ליצור נתונים
4. בסיס מוצק להסקת מסקנות
5. יכולת להסיק סיבה-תוצאה

מחקר תצפיתי

1. צפייה פסיבית
2. תנאים שנוצרו בטבע
3. נתונים ממעקב שותף
4. בסיס חלש להסקת מסקנות
5. חשש שמשתנים מתערבים אחראים לקשרים

The Salk Vaccine- One of the largest experiment in the history

- The polio epidemic – hit USA first in 1916
- By 1950s – several vaccines were discovered
- The one by Jonas Salk *from the University of Pittsburgh* seemed the most promising in laboratories
- In 1954 The Public Health Service decided to do:
 - Large scale field trial
 - Nearly 2,000,000 children involved, among them:
 - Half a million vaccinated
 - One million left deliberately unvaccinated
 - Half a million refused to be vaccinated
 - Field trial among the most vulnerable age group – grades 1,2,3

The Salk Vaccine Trial- Continued

- The 2,000,00 children involved – sample from the population
- Method of analysis - **Comparison:**
 - Treatment is given only to some subjects – Treatment group
 - Others with no treatment (**deliberately unvaccinated**)– Control group
 - Treatment and control are of different sizes, but that is OK
 - What is to be measured: Rates of polio (cases per 100,000) in 2 groups
 - Rates instead of absolute numbers adjust for sample sizes
- Why do we need a control group? Why not give all the vaccine?
 - History: Rates vary from year to year: In 1952-60,000, in 1953-30,000
 - In 1954 rates dropped sharply from 1953, but without control we don't know:
 - **Is it the vaccine?**
 - **Is it no epidemic that year?**
 - The only way to know leave some **deliberately unvaccinated** - controls

The Salk Vaccine Trial - Continued

- How to select the treatments and controls:
 - Obviously, parents' permission is required for vaccination
 - So, one possibility for treatment control:
 - Vaccinate those who consented
 - The others — control group

Is this a good design? — No!

- Known- consent more likely in the higher-income groups
- Those in the lower-income groups are **less** vulnerable to polio
- Paradox, but those in lower-income are more likely to have had natural immunization in childhood — mild form of polio
- So, the two groups **are not comparable**

The Salk Vaccine Trial - Continued

- ⇒ The statistical lesson - *to avoid bias the treatment group and the control group have to be as similar as possible* – except for the treatment
- ⇒ That makes possible to conclude – difference in response – due to treatment and not other effects
- ⇒ When groups differ in other factors – their effect may be *confounded* (mixed up) with that of the treatment
- ⇒ OK, so *no consent* (treatment) vs. *no consent* (control)
- ⇒ So, should we select the treatment and control?
 - **To make a fair comparison -the treatment and control groups have to be selected from *the same population: Children whose parents consented***
 - **Now how to select?**

Slide 17

f1

fuchs, 2/25/2007

The Salk Vaccine Trial - Continued

- ⇒ *Maybe – human judgment to make them “as similar as possible”?*
- ⇒ *Experience show- Substantial bias*
- ⇒ *Solution: RANDOMIZATION*
 - *50% Chance to be in Treatment*
 - *50% to be in control and get PLACEBO*
- ⇒ *PLACEBO – Injection with salt dissolved in water*
- ⇒ *Children (parents) didn't know which group*
- ⇒ *Why? To avoid **Placebo Effect***
- ⇒ *Another precaution: to avoid possible bias from doctors: **DOUBLE BLIND EXPERIMENT***

The Salk Vaccine Trial - Continued

- ⇒ *Summary of Design of Experiment*
- ⇒ **No** to “Give all and Compare years” – Confounded Effect – Epidemic disease – Various natural rates
- ⇒ **No** to “Consent=Treatment, No Consent=Control”
Confounded Effect – Difference in groups – socio-economic and natural prevalence
- ⇒ **Yes** to Definition of Population: **consented**
- ⇒ Assignment to Treatment and Control by RANDOMIZATION
 - PLACEBO – Children (parents) didn’t know which group
 - DOUBLE BLIND EXPERIMENT – Doctors didn’t know

The Salk Vaccine Trial - Continued

- ⇒ *Another of Design of Experiment – Design of National Foundation for Infantile Paralysis (NFIP):*
- ⇒ *Treatment – Control but no Randomization:*
 - *Treatment: Children (consent) grade 2*
 - *Control: Grade 1 and 3*
- ⇒ *Problems:*
 - *Contagious: Rate could be much higher or much lower in grade 2*
 - *Consent in grade 2: bias against vaccine*

The Salk Vaccine Trial – *Results*

<i>Double</i>	<i>Blind</i>	<i>Randomized</i>			
<i>Controlled</i>	<i>Experiment</i>		<i>The</i>	<i>NFIP</i>	<i>Design</i>
	Size	Rate (per100,000)		Size	Rate
Treatment	200,000	28	Grade 2 (vaccine)	225,000	25
Control	200,000	71 (61% reduction)	Grade 1,3 (control)	725,000	54 (54% reduction)
No Consent	350,000	46	Grade 2 (no consent)	125,000	44

Controlled and Uncontrolled Experiment – Gastric Freezing

- ⇒ *Treatment of ulcer – 1958 new technique – gastric freezing: balloon in stomach, coolant pumped and freezes the stomach for 1 hour. The digestive process stops and ulcer begins to heal*
- ⇒ *Wangensteen tried it in 1954 on 24 patients - **all were cured***
- ⇒ ***Good enough proof?***
- ⇒ *Many doctors were suspicious – no controls*
- ⇒ *In 1963 a double blind randomized – controlled experiment was set up by Ruffin:*
 - *82 treatment and 78 control (randomly assigned)*
 - *Treatment – by Wangensteen's technique*
 - *Control – Sham freeze – a shunt into balloon – returned coolant before reaches stomach*
- ⇒ *Two- years follow up by doctors who didn't know whether patients had been given true freeze or sham.*

Controlled and Uncontrolled Experiment – Gastric Freezing - Results

	Treatment	Control
Six-weeks	29% symptom free 47% improved	29% symptom free 39% improved
24 months	45% relapse and clinically worse	40% relapse and clinically worse

At no time during follow-up there was significant difference between the groups ➡

- ➡ Study shows conclusively – freezing procedure no better than sham
- ➡ **Reasonably to assume – early positive reports – psychological effects of the procedure**
- ➡ Wangesteen's experiment – badly designed – no control.
- ➡ Ruffin's experiment – well designed – double blind – randomized controlled

Observational study – Sex Discrimination at Graduate Division at Berkeley (1973)?

	Men	Women
Overall	8,442 applied 44% admitted	4,321 applied 35% admitted
6 largest majors (more than 500 applicant each)	2,961 applied 45% admitted	1,835 applied 30% admitted

- ⇒ *Admissions to graduate work - separately for each major*
- ⇒ *No experiment is possible*
- ⇒ *But we can **control** for the confounding (mixed up) factor*

Observational study—Sex Discrimination—Controlling for Majors

Major	Men-Applicants	Percent Admitted	Women-Applicants	Percent Admitted	
A	825	62%	108	82%	
B	500	63%	25	68%	
C	325	37%	593	34%	
D	417	33%	375	35%	
E	191	28%	393	24%	
F	373	6%	341	7%	

דוגמה 4: תרופה להורדת כולסטרול

ב- 1980 התפרסמו תוצאות מניסוי לבדוק את היעילות של clofibrate לשפר את ההישרדות של חולי לב, תוך הורדת רמת הכולסטרול שלהם.

הניסוי כלל קבוצת טיפול, שקיבל את התרופה, וקבוצת ביקורת, שקיבלה פלסבו – כלומר, תרופת דמי שלא הייתה אמורה להשפיע כלל על בריאותם.

כל החולים קיבלו כדורים שנראו זהים וחולים והרופאים המטפלים לא ידעו מי קיבל את התרופה.

לכן מדובר בניסוי עם כפל סמיות וביקורת פלסבו.

דוגמה 4 (המשך): תוצאות נבחרות

בסוף הניסוי, מגלים לרופאים מי קיבל תרופה ומי קיבל פלסבו ומנתחים את הנתונים.

הם מצאו את התוצאות הבאות לתמותה תוך 5 שנים:

טיפול (תרופה) $n = 1813$ תמותה = 18.2%

ביקורת (פלסבו) $n = 2695$ תמותה = 19.4%

האם התרופה עוזרת או לא כל כך?

דוגמה 4 (המשך): תוצאות נבחרות

יש המשך: החוקרים שמו לב שרבים מן החולים שקיבלו את התרופה לא לקחו את כל הכדורים. כאשר הם חילקו את החולים **שאכן קיבלו את התרופה** על בסיס זה (post-stratification), הם מצאו את התוצאות הבאות לתמותה תוך 5 שנים:

80% או יותר $n = 708$ תמותה = 15.0%

פחות מ- 80% $n = 357$ תמותה = 24.6%

האם התרופה עוזרת רק לאלה שמתמידים לקחת אותה?

דוגמה 4: תוצאות נבחרות

אז בדקו הרופאים מה קרה בקבוצת הפלסבו :

80% או יותר $n = 1813$ תמותה = 15.1%

פחות מ- 80% $n = 882$ תמותה = 28.2%

השימוש בקבוצת פלסבו הראה שההבדל בין הקבוצות נובע, ככל הנראה, מגורמים אחרים – למשל יכולת להתמיד בדיאטה או אורך חיים בריא יותר או טיפול רפואי שוטף יותר – ולא מן התרופה.

דוגמה 5: מציאת מלפפון בעל יבול גבוה

תחנה לניסויים חקלאיים רוצה לערוך השואה בין 5 זנים של מלפפונים כדי לבדוק האם הם שונים מבחינת יבול.

הניסוי יתבצע על שדה ניסויי וכל חלקה בשדה תהיה זרועה באחד הזנים.

בניסוי כזה, יתרון ממוצע של 5% ביבול יכול להיות בעלת חשיבות כלכלית רבה. אבל יש גם פיזור רב – היבול בחלקה בודדת יכולה לסטות מן הממוצע ב- 30% או יותר. לכן, הניסוי צריך להיות מספיק גדול לזהות הבדלים שהרבה מתחת לרמת הפיזור.

מספר סוגיות בתכנון הניסוי

איך כדאי לחלק את השדה לחלקות.

איך לקבוע איזה זן לזרוע בכל חלקה.

האם ניתן ליצור בלוקים של חלקות.

איך לדאוג שהחלקות לכל זן יהיו "ברי השוואה הוגנת".

מהו גודל המדגם הדרוש לניסוי.

דוגמה 6: מציאת מדיום אופטימלי לגידול תאים

חברה פרמאצבטית רוצה לפתח תרופה הדורשת גידול של תאים מסוג מסוים. התאים גדלים בתוך מדיום מועשר אך אין ידע תיאורטי היכול לכוון מה היא ההעשרה האופטימלית.

עובדי החברה הכינו רשימה של כ- 50 תוספים שונים העשויים לשפר את גידול התאים.

יתכן וכדאי להשתמש בכמה תוספים ביחד במדיום.

מספר סוגיות בתכנון הניסוי

יש מספיק זמן לכמה סבבים של ניסויים. איך כדאי לחלק את המאמץ בין הסבבים השונים.

איזה (וכמה) תוספים ניתן לכלול בניסוי.

איזה הרכבים של תוספים כדאי להפעיל ביחד.

דוגמה 7: חקירת חשיבה על בעיות הנדסה

ידוע שאנשים לפעמים טועים בגירויים קוגניטיביים כיוון שהם שמים לב להיבט לא רלוונטי של הגירוי.

ניסוי זה מציג שני מצולעים והנבדק צריך לסמן איזה מצולע בעל היקף גדול יותר. המצולעים שונים גם מבחינת שטח והבדל זה מספק את ההיבט הלא רלוונטי.

לחוקרים תיאוריה לפיה הזמן הדרוש לתת תשובה יהיה תלוי בהתאמה (או אי-התאמה) בסדר של השטחים וההיקפים.

מספר סוגיות בתכנון הניסוי

איך לקבוע כמה פריטים מקבל כל נבדק.

איך לקבוע כמה נבדקים לכלול.

מה יחידת הניסוי כאן? נבדק? מצולע? תשובה?

איך לקבוע את צורות המצולעים – אולי צורת המצולע מכניסה גורמים נוספים שעשויים להשפיע על תוצאות הניסוי.

התקדמות איטרטיבית

לשקול סבב של ניסויים במקום ניסוי אחד כוללני.

