



## Простите промени в индивидуалните изследвания могат да подобрят възпроизводимостта на биомедицинския научен процес

*Независимо от резултата научните изследвания винаги трябва да се публикуват. Това е едно от заключенията на изследователски проект, който е проведен от Германския център за защита на лабораторните животни в Германския федерален институт за оценка на риска (BfR). Проучванията показват, че научните изследвания имат по-голям шанс да бъдат публикувани, ако имат желанието "положителен" резултат, като например измерване на очаквания ефект, откриване на вещество или валидиране на хипотеза. Отрицателните или невалидните резултати, които нямат такъв ефект, имат по-малък шанс за публикуване. Учените също имат голям интерес да постигнат значими резултати, които са достойни за публикуване, като по този начин стимулират научните изследвания. Резултатът от това обаче може да доведе до това, че са публикувани проучвания, резултатите от които не са възпроизводими и поради това изглеждат само "положителни". Тези привидно положителни резултати след това водят до допълнителни проучвания, които се основават на предполагаемия доказан ефект.*

*Математическият модел, представен в публикацията, показва как може да се прекъсне механизмът на "фалшиви положителни" резултати. Всички проучвания - независимо от получените резултати, трябва да бъдат публикувани, след като са спазени добрите научни практики и по този начин неправилният резултат може да бъде опроверган по-бързо. Това означава, че привидно отрицателен резултат не е недостатък, а по-скоро е в плюс за изследването. **Например тестът върху животни, който не може да докаже ефикасността на ново лекарство, не би бил неуспех в очите на науката, а по-скоро ценен резултат, който предотвратява ненужните последващи изследвания (и допълнителни тестове върху животни) и ускорява разработването на нови терапии.** В дългосрочен план могат да бъдат избегнати ненужни последващи тестове с животни въз основа на фалшиви предположения. В крайна сметка използването на повече животни в един експеримент може да намали общия брой на използваните животни.*

В зависимост от мета-изследванията между 51 и 89 процента от резултатите, публикувани от биологични научни изследвания, не могат да бъдат възпроизведени от други изследователи. Невро-научните изследвания показват, че недостатъците в статистическата оценка на експериментите често са причина проучванията да не могат да бъдат възпроизведени.

Разработен е модел за оценка на въздействието на препоръките, коригиращи кризата на възпроизводимостта (чрез публикуване както на положителни, така и на "отрицателни" резултати и увеличаване на статистическата сила) върху конкурентни

цели като откриване на причинно-следствени връзки. За разлика от последните публикации, този модел определя количественото въздействие на всяко едно предложение не само за индивидуално проучване, но и главно върху връзката и последствията за цялостния научен процес. Доказано е, че експериментите с по-висока мощност могат да спестят ресурси в цялостния изследователски процес, без да генерират излишни фалшиви положителни резултати. Колкото по-добро е качеството на информацията от предварителното проучване и нейната експлоатация, толкова по-вероятно е този благоприятен ефект да се получи. Освен това, се определят количествено неблагоприятните ефекти от пренебрегването на добри практики при проектирането и провеждането на изследванията, базирани на хипотези и пропускането на публикуване на "отрицателни" заключения.

## Въведение

Възпроизводимостта може да бъде определена по редица начини. В едно от последните статии на специализираното научно списание „*PLoS Biology*“ е прието определение за понятието възпроизводимост, което обхваща съществуването и разпространението на една или повече грешки, несъответствия или пропуски, които пречат на възпроизводимостта на резултатите. Авторите оценяват, че възпроизводимостта на публикуваните научни данни варира от 51% до 89%, което се подкрепя от мета-изследванията през последните години. Учените, които отговарят за изследването предполагат, че възпроизводимостта на публикуваните научни изследвания е по-малка от очакваните резултати. По-конкретно 90% от 1 576 учени са се съгласили, че има значителна или слаба възпроизводимост. **Интересен факт е, че публикуваните данни не могат да бъдат възпроизведени, но 31% от респондентите все още не вярват, че публикуваните резултати може да са погрешни.** Липсата на възпроизводимост често влошава правдоподобността на науката и предизвиква сериозни опасения за разхищаването на ресурсите, защото се възпрепятства научният прогрес. Предизвиква се ненужното страдание на експериментални животни, тъй като много от транслационните и предклиничните изследвания в биомедицинските изследвания се основават на експерименти с животни. **Следователно, научната общност се нуждае от решение за справяне с кризата на възпроизводимостта.**

Има много причини, поради които едно проучване не може да бъде възпроизведено. Разбира се, недостатъчното описание на метода води до невъзпроизводими експерименти и данни, което често е оправдано, тъй като разделянето на методите е препратено към малки параграфи в широкоспектърни списания. Следователно методите раздели често не съдържат важна информация за читателите, за да се оцени силата на експерименталните данни и да се направят изводи. Поради това трябва да се подобри отчитането на експерименталния дизайн и методите, за да се избегне възпроизводимостта и устойчивостта на дисбалансите в науката. По-важното е, че дори когато експериментът може да бъде повторен, невъзпроизводимостта се появява, когато предполагаемият ефект (т.е. причинно-следствената връзка) или няма посочената сила, или въобще не съществува. **С други думи, възпроизводимостта се дефинира като способността точно да се повтори експеримент и да се получи идентичен резултат в рамките на допустимите отклонения в статистическата грешка.** От различни учени се идентифицират редица причини, които допринасят за невъзпроизводимостта на повечето публикувани проучвания през 2005 г. Тези причини

понастоящем се обсъждат интензивно от научната общност и са получени няколко практически препоръки за намиране на изход от "кризата на възпроизводимостта".

**Тези препоръки се обобщават, както следва:**

**А.** Положителните резултати от проучванията показват ефект, но отрицателните резултати показват, че няма доказателства за очаквания ефект. Положителната прогнозна стойност (PPV) се определя като очаквания брой на всички истински положителни резултати се разделя на очаквания брой на всички положителни резултати от изследването. **Обобщавайки PPV в науката е мярка за вероятността.** Това количество се увеличава с вероятността от предварителното изследване и намалява с вероятността за грешки в алфа и бета. Неблагоприятната селекция от хипотези (ефекти), които трябва да се изследват, води до ниска степен на вероятност от предварителна подготовка и по този начин намалява PPV (т.е., броят на фалшиво положителните резултати се увеличава).

**В.** Основен проблем, който се разглежда в литературата е "гъвкав дизайн" (т.е. разликите между рандомизирани и наблюдателни опити не се вземат предвид, когато се планира проучване) и "гъвкав анализ на данните" за експерименти (т.е. анализ на едни и същи данни), които увеличават тенденцията (пристрастието) към положителните резултати. Систематичните грешки при статистическата оценка на експериментите в неврологията са една от причините, които биха могли да доведат до отклонение в полза на положителните резултати.

**С.** Вероятността за успешно публикуване на "отрицателни"/нулеви резултати е по-ниска от тази за положителните резултати. Например "негативните"/нулевите резултати многократно остават непрочетени в редица лабораторни книги и не се публикуват, защото те са отхвърлени от научните списания. Но ако резултатите от "негативни"/нулеви резултати станат публикувани, тези резултати често се съобщават от списанията с ниска степен на въздействие или като странични продукти с положителни резултати и поради това са трудни за наблюдение. Съответно голям брой резултати от изследванията остават незабелязани. Разглеждането на такива резултати е много по-ниско за по-нататъшни проучвания и по този начин не може да повлияе на научния прогрес. Накратко, липсата на информация за "отрицателните"/нулеви резултати води до отклонение от публикуването, тъй като положителните ефекти са надценени (например в мета-проучвания) и фалшиво-положителните резултати не се опровергават.

**Д.** По-ниската мощност означава, че фалшивият отрицателен процент се увеличава, което води до по-малка вероятност за откриване на истинските ефекти. Освен това, вероятността положителната находка да е вярна (т.е. PPV) също е намалена, тъй като PPV намалява с намаляване на мощността. Поради това се препоръчва предварителен анализ на пробите за постигане на подходяща мощност. Поради това се препоръчват по-малко, но по-големи проучвания, тъй като резултатите от изследванията стават по-вероятни.

Всъщност, всичките четири препоръки звучат много правдоподобно за паралелни експерименти. Въз основа на тези предпоставки може да се приеме, че препоръчаните промени подобряват вярата в науката, измерена от PPV. Няколко модела са публикувани през последните години. Възможно е да се разграничат моделите, които

предсказват поведението на учените в дадена система от стимули и тези модели, които са предназначени да изследват въздействието на това поведение върху ефикасността и качеството на науката. Отправна точка на такова разглеждане е, че съществува разлика между интереса на отделните учени и интереса на обществеността към качеството и ефективността на науката. При опростен подход броят на публикациите е от решаващо значение за перспективата за кариера на учен. *Gall et al.* въвеждат голям набор от методи за икономика и социални науки за разследване на ефектите от стимулиращите системи. Други модели се фокусират върху въздействието на променено поведение, като например използването на по-големи размери на извадките за ефективността и качеството на науката.

*De Winter & Happee u van Assen et al.* разглеждат въздействието на "селективното публикуване" срещу "публикуването на всичко" в метааналитичен процес, при който се изследва един ефект, докато силата на ефекта се определя с достатъчна сигурност.

В конкретния математически модел се разглеждат следният аспект на научния процес:

- Публикуваните експериментални резултати може да причинят или предотвратят по-нататъшни експерименти, свързани с един и същ изследователски проблем, въпреки че се стремят към различни потенциални причинно-следствени връзки (в най-лошия случай: "лошата наука проправя пътя към по-лоша наука").

Въздействието на препоръките трябва да се измерва в три цели за един разширен модел:

- а. разкриване на истинската причинна връзка (чувствителност);
- б. избягване на фалшиви положителни резултати (специфичност);
- с. минимизиране на общото потребление на ресурси.

## Модел

Примерен случай е един единствен изследователски екип, който изследва и наблюдава даден ефект, като например задействане на сигнална каскада от определена малка молекула. Изследвани са две научни хипотези: малката молекула "стимулира рецептора А" или "инхибира рецептора В". В допълнение се предполага, че този конкретен изследователски проблем и неговите резултати са достатъчно интересни или релевантни за научната общност. Например, други изследователски екипи разглеждат "съединения, които стимулират рецептор А" и може да използват "малката молекула" като положителна контрола за своя експеримент.

Приема се, че "малката молекула стимулира рецептора А" и това е първата научна хипотеза. За да се изчисли размера на извадката, необходим за експеримента, екипът решава за ниво на значимост  $\alpha$  и статистическа мощност  $(1-\beta)$ , като се има предвид очакваният размер на ефекта  $\delta$ . След завършване на експеримента статистическият тест оценява нулевата хипотеза. Важно е да се отбележи разликата между научната и статистическата хипотеза. За да се постигне по-добро разграничение, се използва термина "фактор" за научната хипотеза. Оценката на нулевата хипотеза води до нейното отхвърляне ("малка молекула стимулира рецептор А") или приемането му

("малка молекула не стимулира рецептор А"). Освен отхвърлянето на нулевата хипотеза не е синоним на приемане на алтернативната научна хипотеза ("малката молекула инхибира рецептора В"). Ако се предположи, че изследователският екип открива как "малка молекула стимулира рецептора А". Ако един изследователски екип вярва, че научният въпрос трябва да бъде решен, то той не разглежда други проучвания, но и не знае дали резултатът е вярно или фалшиво положителен. Тъй като резултатът от статистическия тест може да бъде или верен или не, има четири различни възможни резултата: **истински положителен, фалшив положителен, истински отрицателен и фалшив отрицателен**. Разпределението на тези резултати зависи от вероятността от статистическите грешки  $\alpha$  и  $\beta$ . Те водят до различни общи резултати за изследователския проблем (отговаряйки на въпроса за механизма, с който малката молекула предизвиква сигнализиращата каскада). Следователно изследователският екип няма да тества втория фактор, който е втората научна хипотеза (малката молекула инхибира рецептора В) и ще публикува резултата, че "малката молекула стимулира рецептора А". Типичен сценарий на науката, защото като цяло всички положителни резултати ще бъдат публикувани. В дългосрочен план тази научна процедура води до важна разлика между публикуването на истински положителен и фалшив положителен резултат.

В друг случай изследователският екип ще продължи да тества своята втора научна хипотеза ("малката молекула инхибира рецептора В"), ако статистическият анализ на първата научна хипотеза показва отрицателен резултат: откриването на отрицателен резултат ще позволи изследователският екип да изпробва втората вероятна хипотеза, за да намери решение на своя изследователски проблем. Публикуването на отрицателния резултат няма да има никакво въздействие, тъй като само този единствен изследователски екип работи по този конкретен научен въпрос. Фалшивият отрицателен резултат в крайна сметка ще доведе до края на изследователския процес (всички налични хипотези са тествани) без откритие, защото истинската първа хипотеза е отхвърлена, а втората хипотеза също ще се отхвърлена (дори и възможно фалшив положителен резултат в крайна сметка ще се разкрие поради изследвания, основани на споменатия резултат (фалшифицирани).

Публикуваните резултати от отделни научни изследвания могат да окажат влияние върху посоката на изследването. Това означава, че фалшивите "отрицателни"/нулеви резултати могат да доведат до ненужни (последващи) експерименти. За моделирането се предполага, че изследователският проблем се решава чрез тестване на няколко алтернативни хипотези и то последователно.

Този разширен модел е използван, за да отговори на два въпроса: 1) дали е възможно да се намали общото потребление на ресурси (например животински животи или пациенти) в рамките на научния процес като цяло чрез увеличаване на размера на извадката (например броя на лабораторните животни) в отделните експерименти и 2) при какви обстоятелства и до каква степен е полезно публикуването на "отрицателни"/нулеви резултати?

За първия въпрос, изчисленията на модела показват, че по-големият размер на извадката за индивидуален експеримент може да доведе до намаляване на общото потребление на ресурси. При експериментите с животни това е еквивалентно на намаляване на общия брой на лабораторните животни. Във връзка с втория въпрос моделът показва, че непубликуването на "отрицателни"/нулеви резултати винаги има

вредно въздействие върху общото потребление на ресурси и води до излишък на фалшиви положителни резултати. Заключението, получено от тези резултати, отново е винаги да се публикуват "негативни"/нулеви резултати.

Идентифицирани са поне две приложения за модела. Първо, резултатите подкрепят по-строг начин за налагане на спазването на добра научна практика (ОСП). Редакторите на научните списания могат да наложат това, като изискат декларация от авторите за адекватно проучване и разкриване на "негативни"/нулеви или по-общии резултати. Второто приложение цели да подобри схемите за стимулиране на изследователите, така че интересите на отделните изследователи да служат за подобряване на науката. По отношение на приложението си, моделът може да бъде комбиниран с този на еволюционните или теоретичните модели. Обобщаването на резултатите води до заключението, че модела има и своите ограничения. Предполага се, че науката се самокоригира в дългосрочен план. Необходими са модели, които използват средствата на икономическите и социални науки, за да изследват влиянието на институционалните промени върху поведението на учените.

Освен това, съществен въпрос в биомедицинската област е колко повторения от предклинични проучвания наистина са необходими, за да бъдем убедени, че могат да започнат транслационни изследвания (например клинични опити). Текущият модел не може да отговори на този въпрос, тъй като се вземат предвид само резултатите от бинарни проучвания. След това могат да бъдат тествани различни критерии за класифициране на резултатите от изследването като приемане или отхвърляне на хипотезата или тяхната неубедителност.

#### **Заключение:**

**Премахването на бариерата за публикуване на негативните резултати от изследванията би било лесен начин за подобряване на науката и даване на стимули на учените за допълнителни публикации.**

#### **Източници:**

[https://www.bfr.bund.de/en/press\\_information/2018/27/science\\_learns\\_from\\_its\\_mistakes\\_to\\_o-205286.html](https://www.bfr.bund.de/en/press_information/2018/27/science_learns_from_its_mistakes_to_o-205286.html)

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0202762>

#### **Изготвил:**

*Боряна Иванова  
младши експерт в Дирекция КРОКЦ*

*Дата: 17.10.2018 г.*