

---

# Rechnen mit bedingten Wahrscheinlichkeiten

---

Dr. Dirk Hasenclever  
IMISE, Leipzig

*[Dirk.Hasenclever@IMISE.uni-Leipzig.de](mailto:Dirk.Hasenclever@IMISE.uni-Leipzig.de)*

# RT Cox theorem

---

## □ Theorem:

- There exists a (~unique) strictly increasing function  $pr$  such that for all propositions  $A, B$  and information  $X$ :

$$0 \leq Pr(A|X) \leq 1$$

$$Pr(A|X) = 0 \quad \text{iff} \quad A \text{ known false given } X$$

$$Pr(A|X) = 1 \quad \text{iff} \quad A \text{ known true given } X$$

$$pr(A \ \& \ B \mid X) = p(B \mid A, X) * pr(A|X)$$

$$pr(\neg A \mid X) = 1 - pr(A \mid X)$$

# Produktregel

---

$$\text{pr}(AB) = \text{pr}(A|B) \cdot \text{pr}(B) = \text{pr}(B|A) \cdot \text{pr}(A)$$

$$\text{pr}(A|B) := \frac{\text{pr}(AB)}{\text{pr}(B)}, \text{pr}(B) > 0, \text{ Definition für Frequentisten}$$

$\text{pr}(A|B) \neq \text{pr}(B|A)$  im Allgemeinen

$$\text{pr}(A|B) = \text{pr}(B|A) \cdot \frac{\text{pr}(A)}{\text{pr}(B)}$$

# Totale Wahrscheinlichkeit

---

Sei von  $A_1, \dots, A_k$  genau ein  $A_i$  wahr, d.h.

a) paarweise exklusiv:  $(A_i \& A_j) = F$  (alle  $i \neq j$ )

b) exhaustiv:  $(A_1 \mid A_2 \mid \dots \mid A_k) = T$

Dann gilt für alle  $B$

$$\text{pr}(B) = \sum_{i=1}^k \text{pr}(B|A_i) \cdot \text{pr}(A_i)$$

# Unabhängigkeit

---

A und B sind **unabhängig**, wenn

$$\text{pr}( A \ \& \ B ) = \text{pr}( A ) \cdot \text{pr}( B )$$

oder äquivalent (via Produktregel)

$$\text{pr}( A \mid B ) = \text{pr}( A )$$

$$\text{pr}( B \mid A ) = \text{pr}( B )$$

Unabhängigkeitsannahmen reduzieren Komplexität und vereinfachen Rechnungen enorm.

Information über A enthält keine Information über B und umgekehrt.

# Zum Beispiel: Biden versus Trump

---

Vier Propositionen

B – Wähler wählt Biden

T – Wähler wählt Trump

U – Wähler wählt vor Ort in der Urne

A – Wähler wählt als Absentee durch Briefwahl oder vorab.

B und T sowie U und A sind paarweise exklusiv:

$$\text{pr}( T ) = 1 - \text{pr}( B ) ; \quad \text{pr}( U ) = 1 - \text{pr}( A )$$

# Zum Beispiel: Biden versus Trump II

---

$$\text{pr}( B ) = \text{pr}( B \mid U ) \cdot \text{pr}( U ) + \text{pr}( B \mid A ) \cdot \text{pr}( A )$$

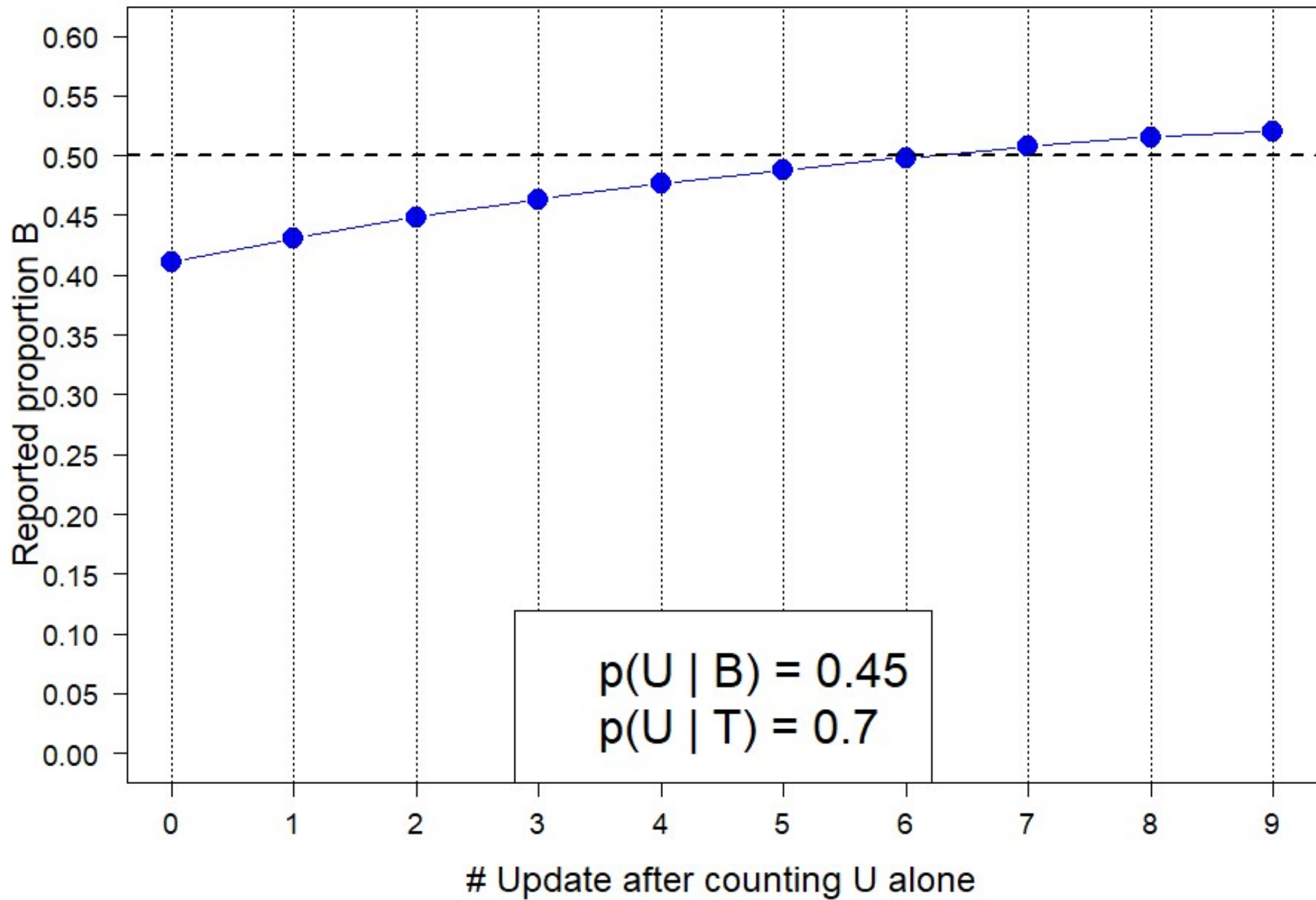
Die Urnen werden zuerst ausgezählt und erst danach die Briefwahl.

Wenn nun B und U **unabhängig wären**:

$$\text{pr}( B \mid U ) = \text{pr}( B ) = \text{pr}( B \mid A )$$

und sollten die Zwischenergebnisse sich nicht relevant ändern.

Da aber  $\text{pr}( B \mid A ) \gg \text{pr}( B \mid U )$   
steigen die Zwischenergebnisse mit fortschreitender  
Auszählung der Briefwahl deutlich an.



# Hotline

---

- Eine Firma beschäftigt drei Mitarbeiter, die telefonische Anfragen von Kunden beantworten sollen.
  - Herr Alleskönner kann 95% aller Frage zur Zufriedenheit der Kunden beantworten,
  - Frau Besserwisser 90% und
  - Herr Chancenlos noch gerade 70%.
- Alle drei Mitarbeiter beantworten gleich viele Telefonate.
- Formalisieren:
  - $\text{pr}( Z | A ) = 0.95$ ;     $\text{pr}( Z | B ) = 0.90$ ;                       $\text{pr}( Z | C ) = 0.70$
  - $\text{pr}( A ) = \text{pr}( B ) = \text{pr}( C ) = 1/3$

# Anwendungen

---

- W. dass ein Kunde mit der Antwort, die er erhält, nicht zufrieden ist.
- $\text{pr}( \neg Z ) = 1 - \text{pr}( Z )$
- $\text{pr}( Z ) = \text{pr}( Z | A ) \cdot \text{pr}( A ) + \text{pr}( Z | B ) \cdot \text{pr}( B ) + \text{pr}( Z | C ) \cdot \text{pr}( C ) = (0.95 + 0.90 + 0.7) / 3 = 0.85$

# Anwendungen

---

- W. dass ein unzufriedener Kunde an Frau Besserwisser geraten ist

$$\begin{aligned}\text{pr}( B \mid \neg Z ) &= \text{pr}(\neg Z \mid B ) \cdot \text{pr}( B ) / \text{pr}(\neg Z ) \\ &= 0.1 \cdot \frac{1}{3} / 0.15 \approx 0.222\end{aligned}$$

# Covid or HIV Tests

---

Patienten haben die Krankheit oder nicht  $D \mid \neg D$   
Test ist positiv oder negativ  $T+ \mid T-$

Eigenschaften des Test**verfahrens**:

Sensitivität:  $\text{pr}(T+ \mid D) = 99.7\%$

Spezifizität:  $\text{pr}(T- \mid \neg D) = 97.0\%$

Aber im Einzelfall interessiert:

Positiv Prädiktiver Wert:  $\text{PPV} = \text{pr}(D \mid T+)$

Negative Prädiktiver Wert:  $\text{NPV} = \text{pr}(\neg D \mid T-)$

# Covid or HIV Tests

---

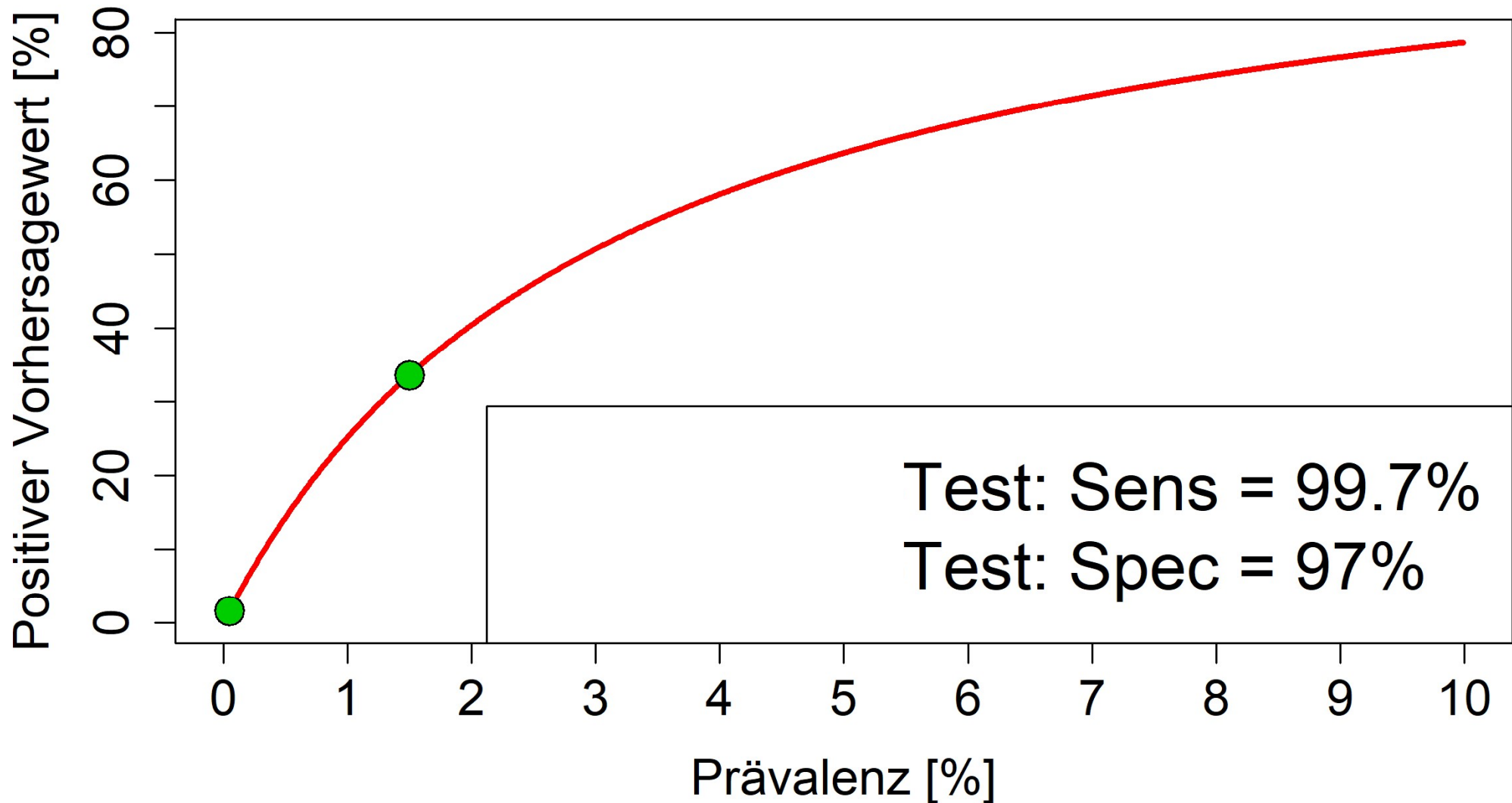
$$\text{pr}( D | T+ ) = \text{pr}( T+ | D ) \cdot \text{pr}( D ) / \text{pr}( T+ )$$

$$= \text{pr}( T+ | D ) \cdot \text{pr}( D ) / \\ ( \text{pr}( T+ | D ) \cdot \text{pr}( D ) + \text{pr}( T+ | \neg D ) \cdot \text{pr}( \neg D ) )$$

$$= \text{pr}( T+ | D ) \cdot \text{pr}( D ) / \\ ( \text{pr}( T+ | D ) \cdot \text{pr}( D ) + (1 - \text{pr}( T- | \neg D ) ) \cdot \text{pr}( \neg D ) )$$

$$= \text{Sensitivität} \cdot \text{Prävalenz} / \\ ( \text{Sensitivität} \cdot \text{Prävalenz} + (1 - \text{Spezifizität}) (1 - \text{Prävalenz}) )$$

## Positiver Vorhersagewert haengt stark von der Prävalenz ab!



# Prävalenzabhängigkeit der Testaussage

## HIV-Test (Elisa)

- Sensitivität: 99.7 % (Anteil Testpositiver von den Infizierten)
- Spezifität: 97.0 % (Anteil der Testnegativen von den Nicht-infizierten)

Prävalenz (Normalbevölkerung) : 0.05%      Prävalenz (in einem diagnostische Labor): 1.5%

|     | T + | T -    |        |
|-----|-----|--------|--------|
| D + | 10  | 0      | 10     |
| D - | 600 | 19 390 | 19 990 |
|     | 610 | 19 390 | 20 000 |

pos. Vorhersagewert:

$$P(D+ | T+) = 10 / 610 = 0.016 = 1.6\%$$

|     | T + | T -    |        |
|-----|-----|--------|--------|
| D + | 299 | 1      | 300    |
| D - | 591 | 19 109 | 19 700 |
|     | 890 | 19 110 | 20 000 |

pos. Vorhersagewert:

$$P(D+ | T+) = 299 / 890 = 0.34 = 34\%$$

# Fast tests for covid-19 are coming

---

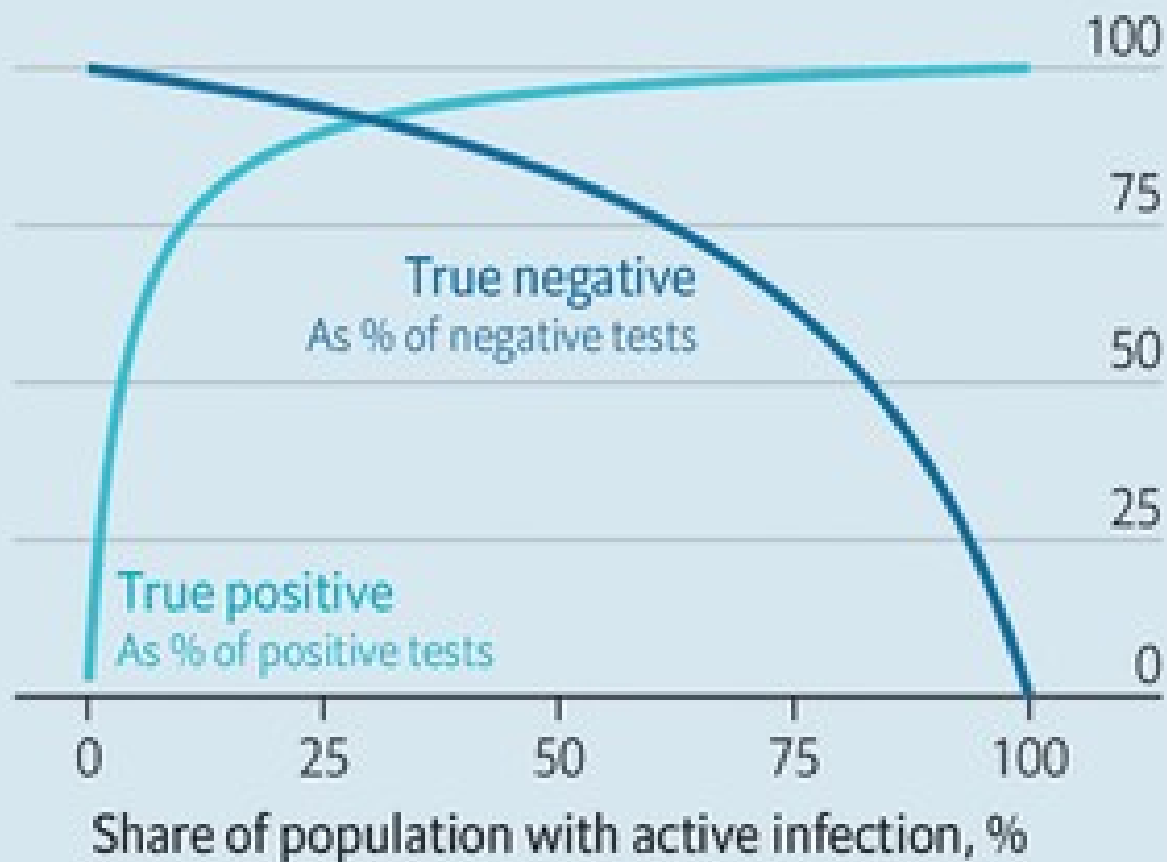
They will help, until a vaccine can be deployed

Rapid tests, by contrast, are designed to detect certain proteins that SARS-COV-2 sheds when it replicates during an infection. These proteins, known as antigens, spur the immune system into making other proteins, called antibodies, that go on to disable the virus. Antigen tests need no laboratory backup and can report a result in 15-20 minutes. They work by dipping the swab into a vial containing a solution that extracts the antigen of interest. A few drops of the mix are then applied to a test strip laced with antibodies that recognise that antigen. The test strip displays the results like a home pregnancy test.

In theory, then, all of this sounds great. But reality is messier. Even a highly accurate test will produce fewer true positives than false positives if the people being tested are unlikely to be infected in the first place (see chart). That would be the kind of problem which arises with mass testing in places that are not covid-19 hotspots. For example, Britain's Office for National Statistics estimates that on October 28th 0.82% of people in private households in London were infected. If everyone in London that day was given a test that has the minimum "acceptable" accuracy for rapid tests set by the WHO (80% sensitivity and 97% specificity) the number of those with false-positive results will be 353% bigger than those with true positive results.

## What is truth?

Covid-19 test results\*, %



Source: *The Economist*

\*For a test with 80% sensitivity, 97% specificity

This is why deciding whether to trust the result of an imperfect rapid test—or, indeed, whether it is worth using the test at all—depends on who is being tested, and why. A positive result is more credible for someone with symptoms, or who is a close contact of an infected individual, and perhaps lives in an area with a high covid-19 rate. But testing people when there is no obvious reason to believe they may be infected is likely to be a waste. A positive result in that case will be suspect.

# Ziegenproblem

---

- In einer Quizshow stehen Sie vor drei identischen Garagentüren.
- Hinter einer der dreien steht ein **roter Ferrari** für Sie, wenn Sie die richtige Tür öffnen.
- Sie stellen sich vor vorläufig vor die Tür1.
- Der Moderator öffnet daraufhin die Tür3:  
**Ätsch eine Ziege!**

Und bietet Ihnen an, sich eventuell umzuentscheiden **!?!**

- **Was tun Sie und warum?**

# Ziegenproblem

---

**Prior:** Wissensstand bevor der Moderator eine andere Tür öffnen.

$$\text{pr}(A1) = \text{pr}(A2) = \text{pr}(A3) = 1/3$$

$$\text{pr}(M3) = 1/2 \quad (\text{eine Tür von den beiden anderen Türen,} \\ \text{Gegeben, dass wir die zuerst gewählte Tür Tür1 nennen.)}$$

**Posterior:** Wissensstand nachdem der Moderator M3 geöffnet hat  
(angenommen der Moderator weiß, wo das Auto steht ist M3 informativ.)

$$\text{pr}(A1 | M3) = \text{pr}(M3 | A1) \text{pr}(A1) / \text{pr}(M3) = 0.5 * 1/3 / 0.5 = 1/3$$

$$\text{pr}(A2 | M3) = \text{pr}(M3 | A2) \text{pr}(A2) / \text{pr}(M3) = 1 * 1/3 / 0.5 = \mathbf{2/3}$$

$$\text{pr}(A3 | M3) = \text{pr}(M3 | A3) \text{pr}(A3) / \text{pr}(M3) = 0 * 1/3 / 0.5 = 0$$

### **Pfizer und BioNTech geben erfolgreiche erste Zwischenanalyse ihres COVID-19-Impfstoffkandidaten in laufender Phase-3-Studie bekannt**

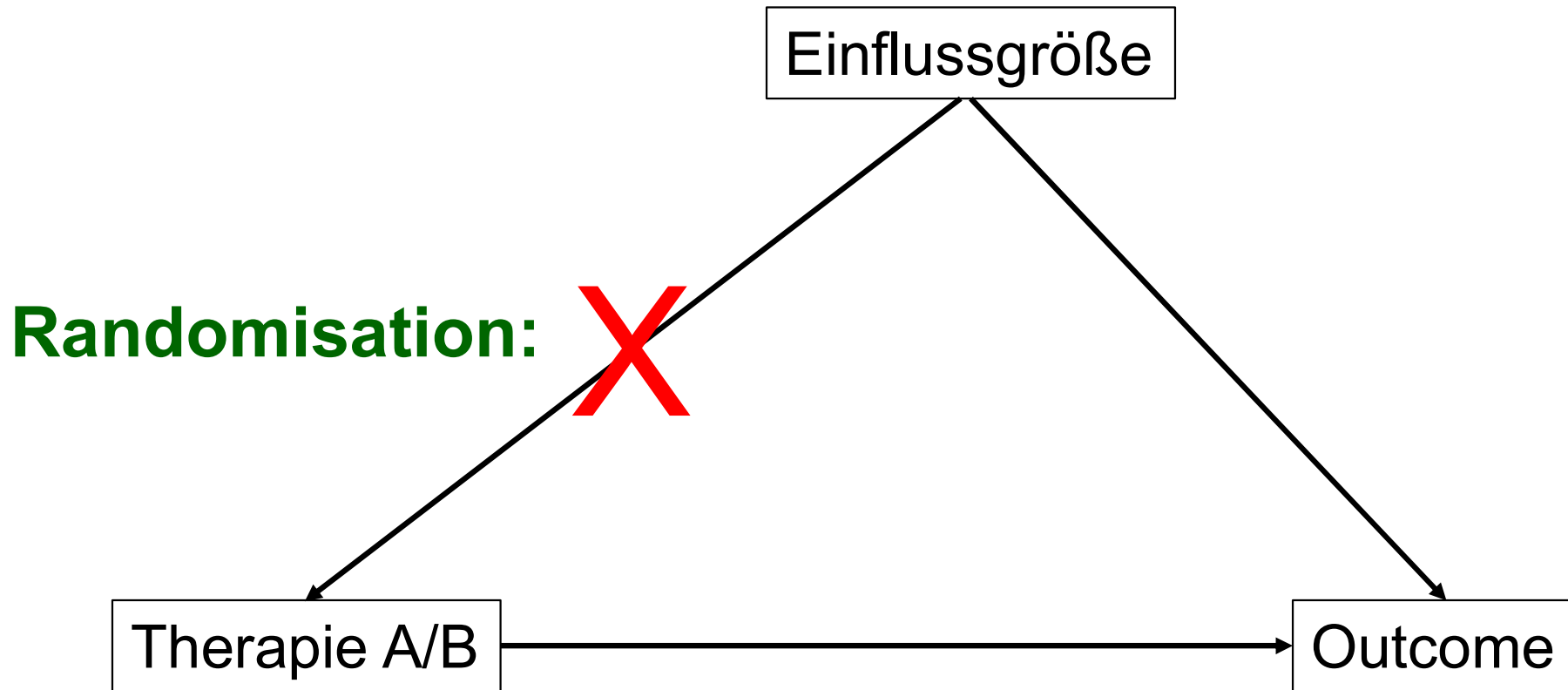
November 9, 2020

- *Impfstoffkandidat erwies sich in der ersten Zwischenanalyse als mehr als 90 % wirksam im Schutz vor COVID-19-Erkrankung in Probanden ohne nachweisliche vorangegangene SARS-CoV-2-Infektion*
- *In der Analyse wurden 94 bestätigte COVID-19-Fälle in Studienteilnehmern untersucht*
- *Insgesamt wurden in die Studie bis heute 43.538 Probanden eingeschlossen. Es wurden keine schwerwiegenden Nebenwirkungen beobachtet, weitere Daten zur Sicherheit und Wirksamkeit werden fortlaufend erhoben*
- *Die Einreichung eines Antrags zur Notfallzulassung (Emergency Use Authorization - EUA) bei der U.S.-amerikanischen Arzneimittelbehörde FDA (Food and Drug Administration) ist kurz nach dem Erreichen des erforderlichen Meilensteins zur Sicherheit geplant. Dieser Meilenstein wird in der dritten Novemberwoche erwartet*
- *Die klinische Studie wird bis zur finalen Datenanalyse mit 164 bestätigten COVID-19-Fällen weiterlaufen, um zusätzliche Daten zu erheben und den Impfstoff in Bezug auf die weiteren Studienendpunkte zu analysieren*

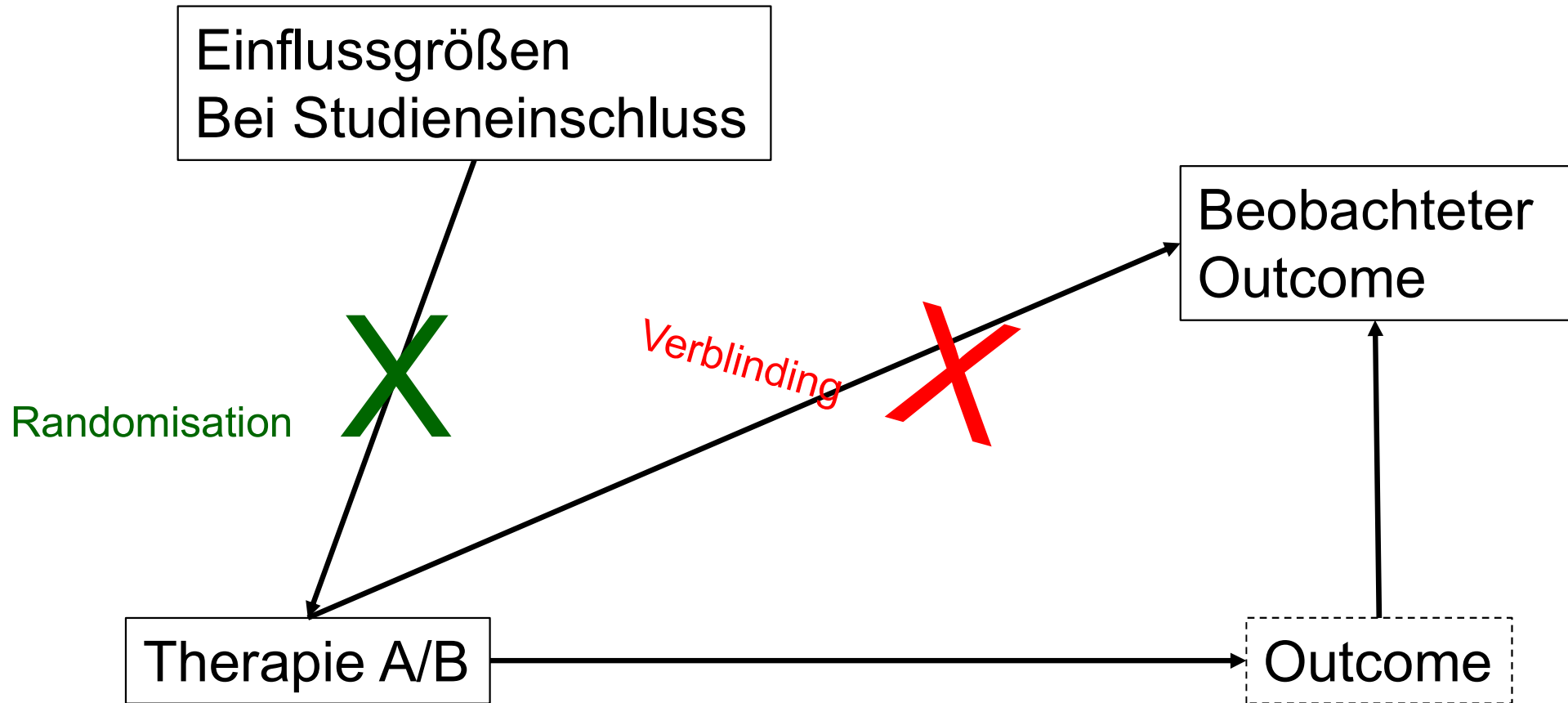
Die Phase-3-Studie zu BNT162b2 hat am 27. Juli begonnen und bis heute 43.538 Probanden rekrutiert. 38.955 der Probanden haben bis zum 8. November bereits die zweite Dosis des Impfstoffs erhalten. Die Studie wird weitere Probanden einschließen und wie geplant fortgesetzt, bis insgesamt 164 bestätigte COVID-19-Fälle aufgetreten sind. Des Weiteren wird die Studie das Potenzial des Impfstoffkandidaten untersuchen, Personen zu schützen, die bereits mit SARS-CoV-2 in Kontakt gekommen sind. Außerdem soll untersucht werden, ob der Impfstoff vor einem schweren Verlauf der COVID-19-Erkrankung schützen kann. Neben dem primären Ziel, die Wirksamkeit anhand bestätigter COVID-19-Fälle zu untersuchen, die sieben Tage nach der zweiten Immunisierung auftraten, wird die finale Analyse, mit Bestätigung der FDA, nun auch die Wirksamkeit des Impfstoffs anhand von bestätigten COVID-19-Fällen analysieren, die 14 Tage nach der zweiten Immunisierung auftraten. Die beiden Unternehmen sind der Ansicht, dass dies die Vergleichbarkeit der Daten zwischen den Studien mit verschiedenen COVID-19-Impfstoffen sowie unterschiedlichen Technologieplattformen verbessern wird und so zu Erkenntnissen aus Quervergleichen beiträgt. Die Unternehmen haben das aktualisierte Studienprotokoll unter folgendem Link veröffentlicht: <https://www.pfizer.com/science/coronavirus>.

# Therapiewahl zufällig und damit unabhängig von allen anderen Einflüssen

---



# Bias nach der Therapieentscheidung



# Endpoint

---

**Was heißt >90% wirksam??**

# Endpoint – Was heißt >90% wirksam??

## 3.2. For Phase 2/3

| Objectives <sup>a</sup>                                                                                                                                                                   | Estimands                                                                                                                                                                                                                       | Endpoints                                                                                                                                                                                                                                            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Primary Efficacy</b>                                                                                                                                                                   |                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                      |
| To evaluate the efficacy of prophylactic BNT162b2 against confirmed COVID-19 occurring from 7 days after the second dose in participants without evidence of infection before vaccination | In participants complying with the key protocol criteria (evaluable participants) at least 7 days after receipt of the second dose of study intervention:<br>$100 \times (1 - \text{IRR})$ [ratio of active vaccine to placebo] | COVID-19 incidence per 1000 person-years of follow-up based on central laboratory or locally confirmed NAAT in participants with no serological or virological evidence (up to 7 days after receipt of the second dose) of past SARS-CoV-2 infection |

IRR = Incidence Rate Ratio ;

Incidence Rate = #events / person time at risk

# Wie stark ist die Evidenz?

---

I

Genaue Zahlen sind noch nicht bekannt, aber ein mögliches Szenario ist:

Von 94 COVID Fällen hatten bei Entblindung durch das DSMC

- 8 den Impfstoff und
- 86 das Placebo

$IRR \approx 100 * (1 - 8/86) = 90.7\%$  (bei Annahme von gleicher person time at risk)