UNIVERSITÄT BERN

Warum randomisieren wir nochmal?

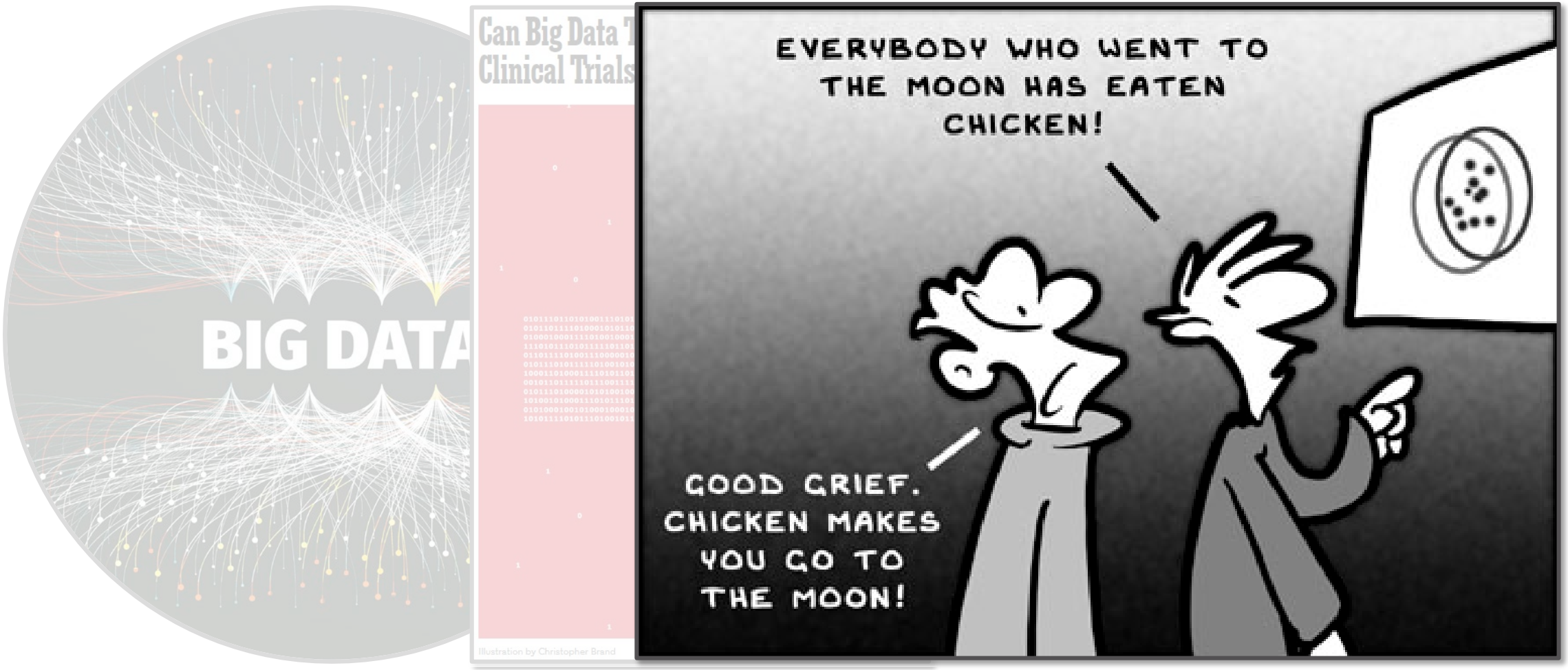
Sven Trelle
CTU Bern
University of Bern

11.9.2018, CTU Lecture

Behandlungsentscheidungen

Ursache-Wirkung

Big data


BIG DATA
Can Big Data T
Clinical Trials
Illustration by Christopher Brand
EVERYBODY WHO WENT TO THE MOON HAS EATEN CHICKEN!
GOOD GRIEF. CHICKEN MAKES YOU GO TO THE MOON!

Individuelle Wirkungsbeziehungen

Kontrafaktische (counterfactual) Wirkungen

Table 2 Counterfactual outcomes of subjects in a study with dichotomous exposure *A* and outcome *Y*

ID	$Y_{a=0}$	$Y_{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Circe	0	1
Ares	1	1
Athene	1	1
Eros	0	1
Aphrodite	0	1
Prometheus	0	1
Selene	1	1
Hermes	1	0
Eos	1	0
Helios	1	0

Hernan MA & Robins JM 2015

Table 2 Counterfactual outcomes of subjects in a study with dichotomous exposure *A* and outcome *Y*

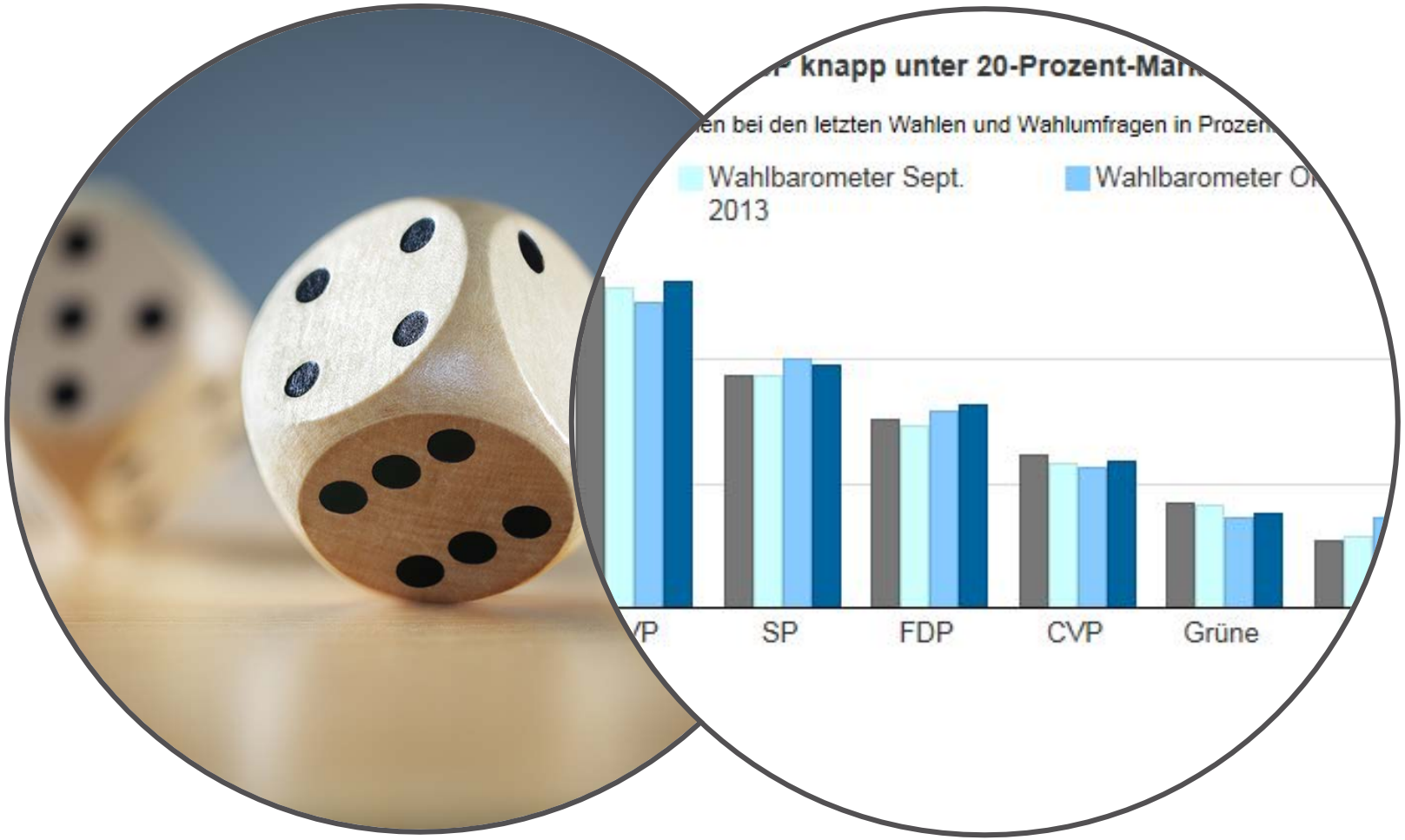
ID	$Y_{a=0}$	$Y_{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Circe	0	1
Ares	1	1
Athene	1	1
Eros	0	1
Aphrodite	0	1
Prometheus	0	1
Selene	1	1
Hermes	1	0
Eos	1	0
Helios	1	0

Kontrafaktische (counterfactual) Wirkungen

- Konzeptionell (theoretisch) definierbar
- Nicht beobachtbar, bestimmbar, messbar
  - Experimentell mit bestimmten Annahmen (Ende)

Wirkungsbeziehungen in Populationen

Zufallsstichprobe



Wirkungsbeziehungen

Voraussetzungen

- Saubere Randomisierung
  - Zufallsmechanismus
  - Verdeckte Zuteilung (Concealment of Allocation)
- Post-Randomisierungs Confounding
  - Co-Interventionen (Performance Bias)
  - Adherence
  - Endpunkterhebung (Detection Bias)
- Differential Losses to Follow-up

Klarer Behandlungsstart

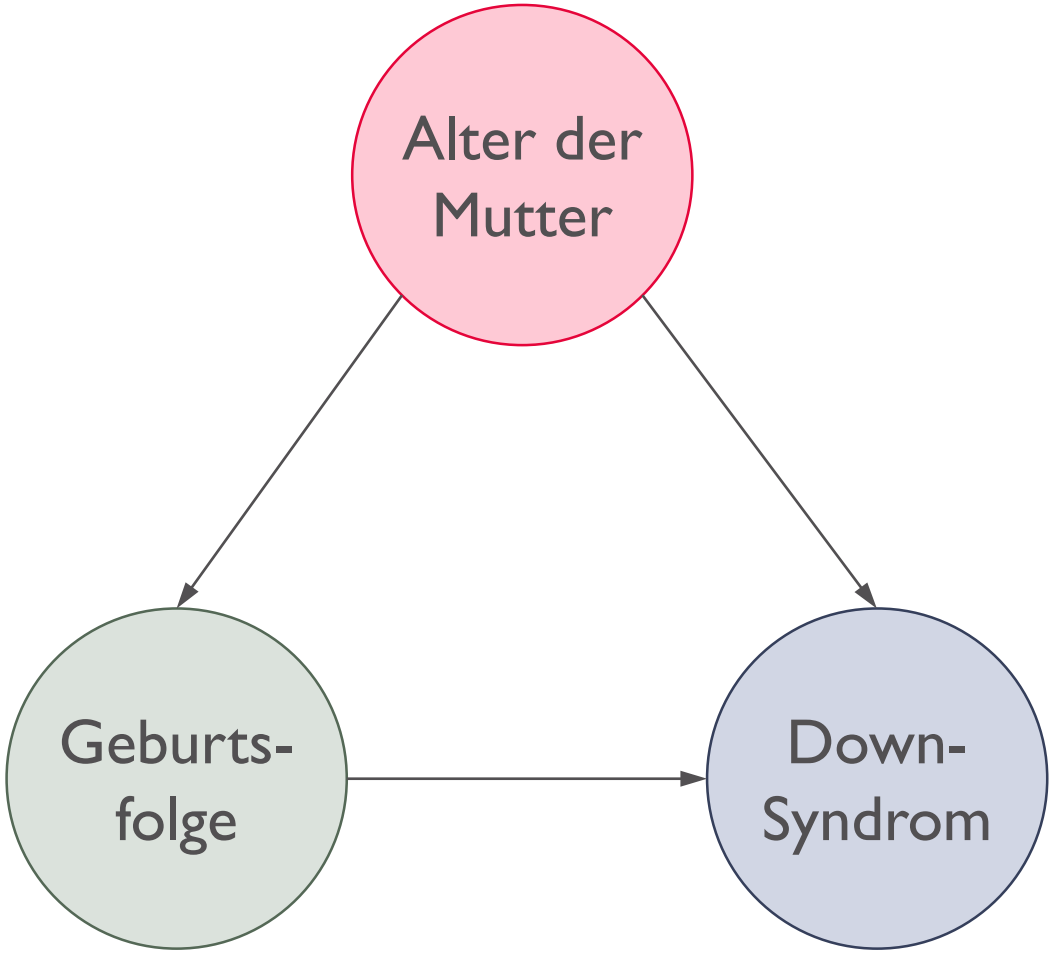
Table 2. **Risk for Major Coronary Heart Disease among Current Postmenopausal Hormone Users and Nonusers, Nurses' Health Study, 1976–1996**

Hormone Use	Person-Years of Follow-up	Cases, *n*	Age-Adjusted Relative Risk (95% CI)	Multivariate-Adjusted Relative Risk (95% CI)*
Never	358 125	662	1.0 (referent)	1.0 (referent)
Past	185 497	337	0.88 (0.77–1.00)	0.82 (0.72–0.94)
Current	265 203	259	0.54 (0.46–0.62)	0.61 (0.52–0.71)
<1 y†	20 091	9	0.30 (0.16–0.58)	0.40 (0.21–0.77)
1–1.9 y†	19 155	9	0.32 (0.16–0.61)	0.41 (0.21–0.80)
2–4.9 y†	78 928	60	0.47 (0.36–0.61)	0.53 (0.41–0.70)
5–9.9 y†	77 435	74	0.51 (0.40–0.65)	0.58 (0.45–0.74)
≥10 y†	69 594	107	0.69 (0.56–0.85)	0.74 (0.59–0.91)

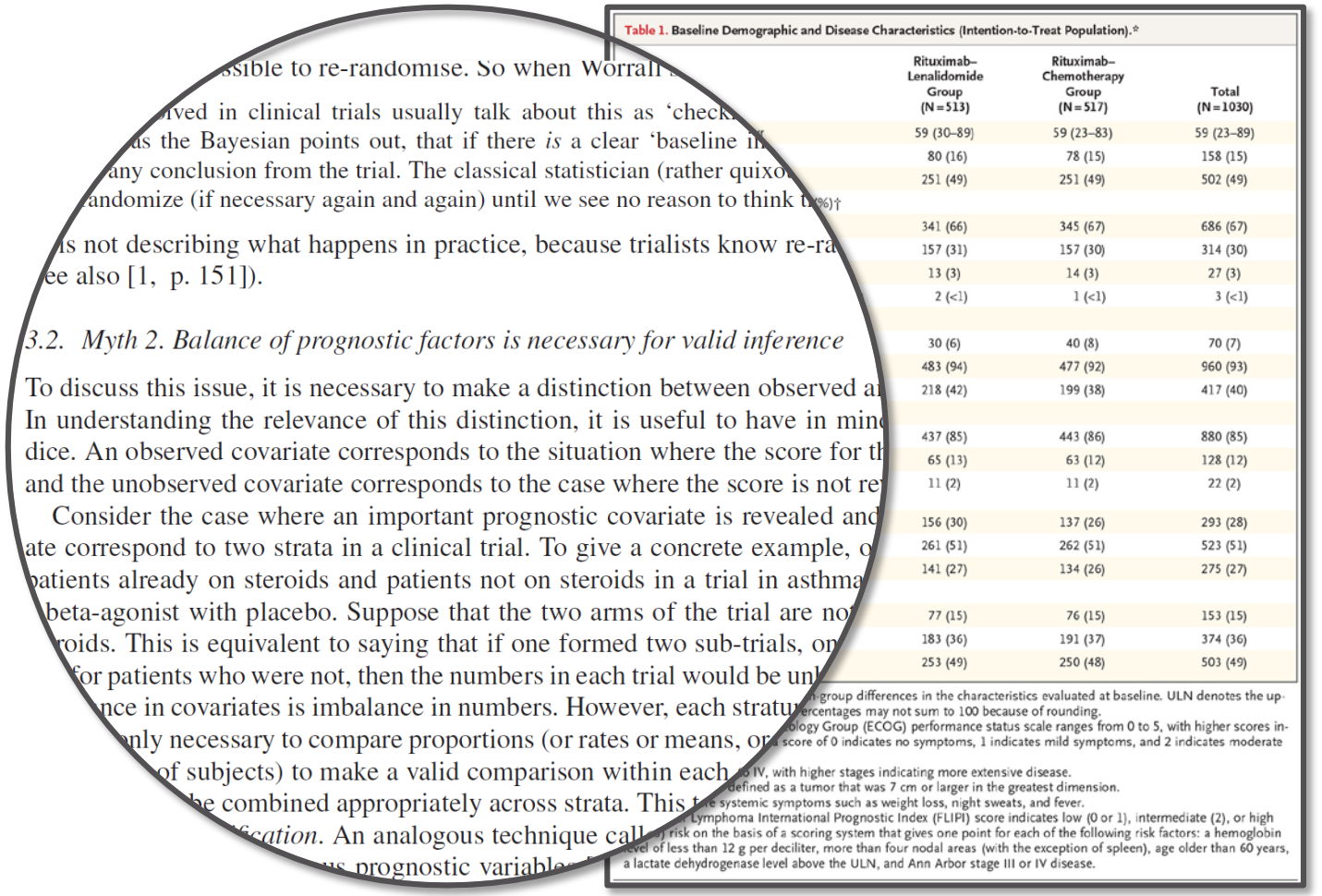
* Adjusted for age, body mass index, history of diabetes, hypertension, high cholesterol level, age at menopause, cigarette smoking, and parental history of premature heart disease.
† Duration of use is underestimated by an average of 1 year, since duration during each 2-year follow-up period was established at the start of each period.

Grodstein F et al. 2000

Confounding?



Müssen prognostische Faktoren und Confounder ausgeglichen sein?



ssible to re-randomise. So when Worrall

ived in clinical trials usually talk about this as 'check
as the Bayesian points out, that if there *is* a clear 'baseline i
any conclusion from the trial. The classical statistician (rather quixo
andomize (if necessary again and again) until we see no reason to think t

is not describing what happens in practice, because trialists know re-ra
ee also [1, p. 151]).

3.2. Myth 2. Balance of prognostic factors is necessary for valid inference

To discuss this issue, it is necessary to make a distinction between observed a
In understanding the relevance of this distinction, it is useful to have in min
dice. An observed covariate corresponds to the situation where the score for th
and the unobserved covariate corresponds to the case where the score is not re

Consider the case where an important prognostic covariate is revealed and
ate correspond to two strata in a clinical trial. To give a concrete example, o
patients already on steroids and patients not on steroids in a trial in asthma
beta-agonist with placebo. Suppose that the two arms of the trial are no
roids. This is equivalent to saying that if one formed two sub-trials, on
for patients who were not, then the numbers in each trial would be unb
nce in covariates is imbalance in numbers. However, each stratu
nly necessary to compare proportions (or rates or means, or
of subjects) to make a valid comparison within each
be combined appropriately across strata. This
fication. An analogous technique call
prognostic variable

Table 1. Baseline Demographic and Disease Characteristics (Intention-to-Treat Population).*

	Rituximab–Lenalidomide Group (N=513)	Rituximab–Chemotherapy Group (N=517)	Total (N=1030)
	59 (30–89)	59 (23–83)	59 (23–89)
	80 (16)	78 (15)	158 (15)
	251 (49)	251 (49)	502 (49)
%)†			
	341 (66)	345 (67)	686 (67)
	157 (31)	157 (30)	314 (30)
	13 (3)	14 (3)	27 (3)
	2 (<1)	1 (<1)	3 (<1)
	30 (6)	40 (8)	70 (7)
	483 (94)	477 (92)	960 (93)
	218 (42)	199 (38)	417 (40)
	437 (85)	443 (86)	880 (85)
	65 (13)	63 (12)	128 (12)
	11 (2)	11 (2)	22 (2)
	156 (30)	137 (26)	293 (28)
	261 (51)	262 (51)	523 (51)
	141 (27)	134 (26)	275 (27)
	77 (15)	76 (15)	153 (15)
	183 (36)	191 (37)	374 (36)
	253 (49)	250 (48)	503 (49)

group differences in the characteristics evaluated at baseline. ULN denotes the up-
ercentages may not sum to 100 because of rounding.
ology Group (ECOG) performance status scale ranges from 0 to 5, with higher scores in-
score of 0 indicates no symptoms, 1 indicates mild symptoms, and 2 indicates moderate

s IV, with higher stages indicating more extensive disease.
defined as a tumor that was 7 cm or larger in the greatest dimension.
e systemic symptoms such as weight loss, night sweats, and fever.
Lymphoma International Prognostic Index (FLIPI) score indicates low (0 or 1), intermediate (2), or high
risk on the basis of a scoring system that gives one point for each of the following risk factors: a hemoglobin
evel of less than 12 g per deciliter, more than four nodal areas (with the exception of spleen), age older than 60 years,
a lactate dehydrogenase level above the ULN, and Ann Arbor stage III or IV disease.

Senn S 2013

Table 1. Baseline Demographic and Disease Characteristics (Intention-to-Treat Population).*

Characteristic	Rituximab–Lenalidomide Group (N=513)	Rituximab–Chemotherapy Group (N=517)	Total (N=1030)
Median age (range) — yr	59 (30–89)	59 (23–83)	59 (23–89)
Age >70 yr — no. (%)	80 (16)	78 (15)	158 (15)
Male sex — no. (%)	251 (49)	251 (49)	502 (49)
ECOG performance status — no. (%)†			
0	341 (66)	345 (67)	686 (67)
1	157 (31)	157 (30)	314 (30)
2	13 (3)	14 (3)	27 (3)
Could not be evaluated or data missing	2 (<1)	1 (<1)	3 (<1)
Ann Arbor stage — no. (%)‡			
I or II	30 (6)	40 (8)	70 (7)
III or IV	483 (94)	477 (92)	960 (93)
Bulky disease — no. (%)§	218 (42)	199 (38)	417 (40)
Follicular lymphoma grade — no. (%)			
1 or 2	437 (85)	443 (86)	880 (85)
3a	65 (13)	63 (12)	128 (12)
Unspecified grade or grade other than 1, 2, or 3a	11 (2)	11 (2)	22 (2)
Lactate dehydrogenase >ULN — no. (%)	156 (30)	137 (26)	293 (28)
$Beta_2$-microglobulin >ULN — no. (%)	261 (51)	262 (51)	523 (51)
B symptoms — no. (%)¶	141 (27)	134 (26)	275 (27)
FLIPI score — no. (%)‖			
0 or 1	77 (15)	76 (15)	153 (15)
2	183 (36)	191 (37)	374 (36)
3 to 5	253 (49)	250 (48)	503 (49)

* There were no significant between-group differences in the characteristics evaluated at baseline. ULN denotes the upper limit of the normal range. Percentages may not sum to 100 because of rounding.
† The Eastern Cooperative Oncology Group (ECOG) performance status scale ranges from 0 to 5, with higher scores indicating greater disability; a score of 0 indicates no symptoms, 1 indicates mild symptoms, and 2 indicates moderate symptoms.
‡ Stages range from I to IV, with higher stages indicating more extensive disease.
§ Bulky disease was defined as a tumor that was 7 cm or larger in the greatest dimension.
¶ B symptoms are systemic symptoms such as weight loss, night sweats, and fever.
‖ A Follicular Lymphoma International Prognostic Index (FLIPI) score indicates low (0 or 1), intermediate (2), or high (3 to 5) risk on the basis of a scoring system that gives one point for each of the following risk factors: a hemoglobin level of less than 12 g per deciliter, more than four nodal areas (with the exception of spleen), age older than 60 years, a lactate dehydrogenase level above the ULN, and Ann Arbor stage III or IV disease.

Morschhauser F et al. 2018

Je grösser die Studie umso weniger Imbalance?

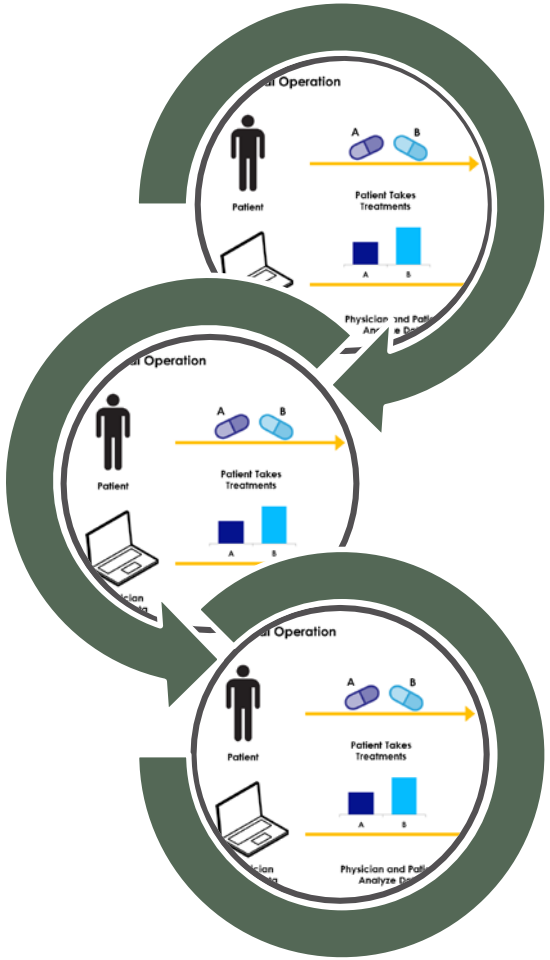
Table 1. Baseline Demographic and Disease Characteristics (Intention-to-Treat Population).*

Characteristic	Rituximab–Lenalidomide Group (N=513)	Rituximab–Chemotherapy Group (N=517)	Total (N=1030)
Median age (range) — yr	59 (30–89)	59 (23–83)	59 (23–89)
Age >70 yr — no. (%)	80 (16)	78 (15)	158 (15)
Male sex — no. (%)	251 (49)	251 (49)	502 (49)
ECOG performance status — no. (%)†			
0	341 (66)	345 (67)	686 (67)
1	157 (31)	157 (30)	314 (30)
2	13 (3)	14 (3)	27 (3)
Could not be evaluated or data missing	2 (<1)	1 (<1)	3 (<1)
Ann Arbor stage — no. (%)‡			
I or II	30 (6)	40 (8)	70 (7)
III or IV	483 (94)	477 (92)	960 (93)
Bulky disease — no. (%)§	218 (42)	199 (38)	417 (40)
Follicular lymphoma grade — no. (%)			
1 or 2	437 (85)	443 (86)	880 (85)
3a	65 (13)	63 (12)	128 (12)
Unspecified grade or grade other than 1, 2, or 3a	11 (2)	11 (2)	22 (2)
Lactate dehydrogenase >ULN — no. (%)	156 (30)	137 (26)	293 (28)
$Beta_2$-microglobulin >ULN — no. (%)	261 (51)	262 (51)	523 (51)
B symptoms — no. (%)¶	141 (27)	134 (26)	275 (27)
FLIPI score — no. (%)‖			
0 or 1	77 (15)	76 (15)	153 (15)
2	183 (36)	191 (37)	374 (36)
3 to 5	253 (49)	250 (48)	503 (49)

* There were no significant between-group differences in the characteristics evaluated at baseline. ULN denotes the upper limit of the normal range. Percentages may not sum to 100 because of rounding.
† The Eastern Cooperative Oncology Group (ECOG) performance status scale ranges from 0 to 5, with higher scores indicating greater disability; a score of 0 indicates no symptoms, 1 indicates mild symptoms, and 2 indicates moderate symptoms.
‡ Stages range from I to IV, with higher stages indicating more extensive disease.
§ Bulky disease was defined as a tumor that was 7 cm or larger in the greatest dimension.
¶ B symptoms are systemic symptoms such as weight loss, night sweats, and fever.
‖ A Follicular Lymphoma International Prognostic Index (FLIPI) score indicates low (0 or 1), intermediate (2), or high (3 to 5) risk on the basis of a scoring system that gives one point for each of the following risk factors: a hemoglobin level of less than 12 g per deciliter, more than four nodal areas (with the exception of spleen), age older than 60 years, a lactate dehydrogenase level above the ULN, and Ann Arbor stage III or IV disease.

Morschhauser F et al. 2018

Individuelle Wirkungsbeziehungen



- Wirkungsbeziehungen untersuchen
  - —Austauschbarkeit der Gruppen
  - —Unterschiede zufällig
- Klarer Behandlungsstart
- Voraussetzung für Anwendung von statistischen Standardverfahren

Vielen Dank!

- Greenland S. Epidemiology 1990; 1: 421
- Hernan MA, Robins JM. Causal Inference 2015
- Rothman K. J Gen Intern Med 2014; 29: 1060
- Senn S. Stat Med 2013; 32: 1439