Big Data for Randomized Controlled Trials:

Opportunities & challenges for the reliable assessment of treatment effects

Martin Landray
Professor of Medicine & Epidemiology
Clinical Trial Service Unit and Big Data Institute
University of Oxford

martin.landray@ndph.ox.ac.uk

UNIVERSITY OF OXFORD

Big Data for assessing clinical outcomes

- More efficient collection of traditional data
  - clinical outcomes (e.g. registries, primary & secondary care)
  - symptoms / quality of life (e.g. NYHA class, EQ5D)
  - economic & social consequences (e.g. education, return to work)
- Novel assessment of traditional disease features
  - exercise capacity (e.g. accelerometer, GPS, wifi)
  - cognitive function (e.g. interactive games)
  - extended physiological assessments (e.g. camera-based BP, respiratory rate, O_2 saturation)
- Novel assessment of new endpoints
  - motor function (e.g. tremor sensor, keystroke speed)

Big Data for reliable evaluation of treatment effects

Scale:	Large number of participants & outcomes Good power for moderate treatment effects
Breadth:	Diverse population (e.g. disease, treatment) Comprehensive safety & efficacy assessments
Length:	Frequency & duration of observations
Depth:	Careful characterization of participants Appropriately detailed classification of outcomes

Enhanced ability to assess impact of health interventions on traditional and novel endpoints

BUT

Fundamental principles of large randomized trials unaltered

Impact of errors on the reliability of results

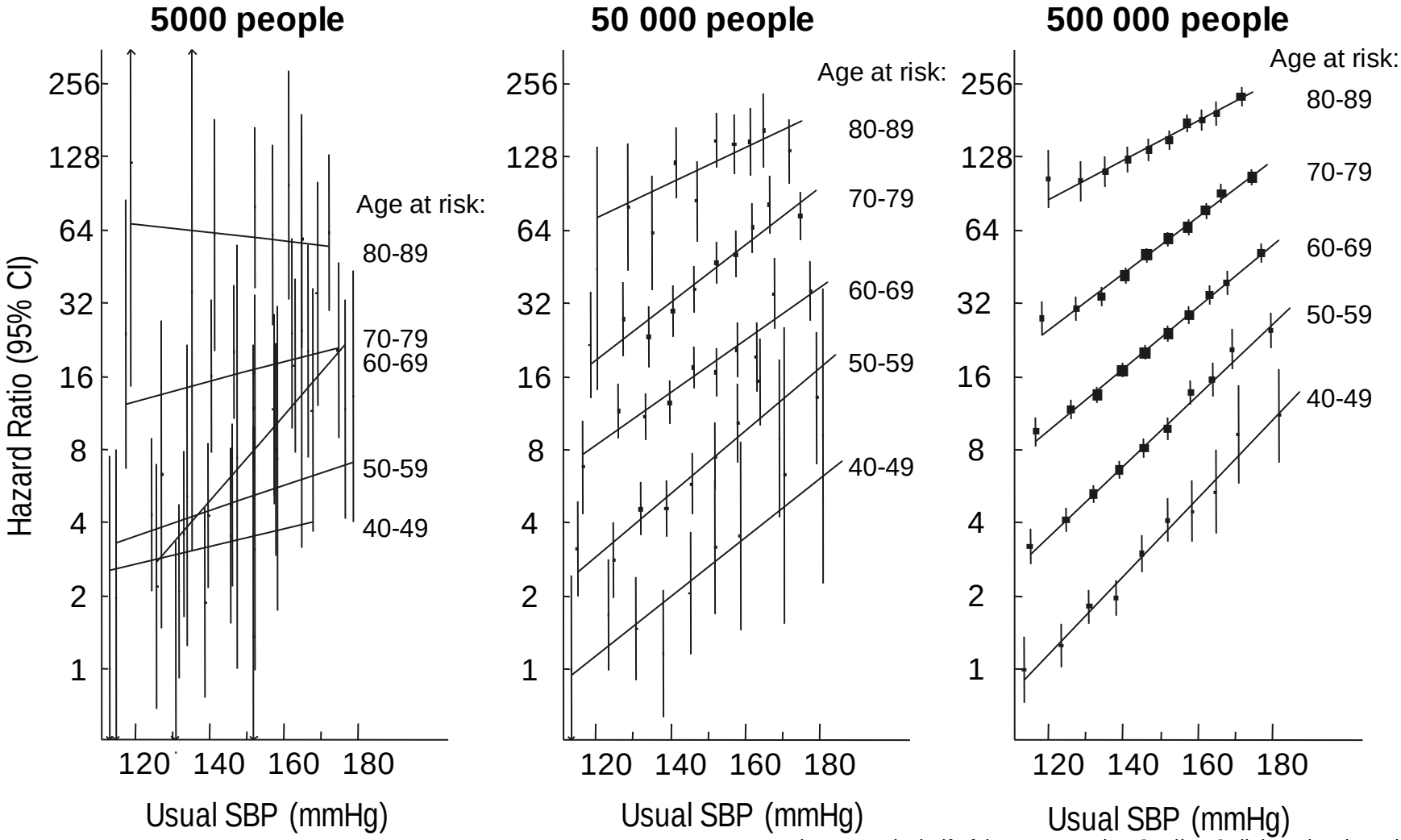
Accurate DATA ≠ Reliable RESULT

- **Random Errors**
  - add noise -> reduces power -> minimizes a difference
  - does not bias the result in any direction
- **Systematic Errors**
  - add bias -> lead towards a particular decision
  - direction & extent difficult to assess

Large *randomized* trials (appropriately analysed) are resistant to small random errors in the data

Data do not need to be perfect

Large numbers give clarity:
Ischaemic heart diseases vs. Systolic BP



Lewington on behalf of the Prospective Studies Collaboration (unpublished)

Outcome ascertainment & adjudication:
Minimal impact of including false events / missing real events

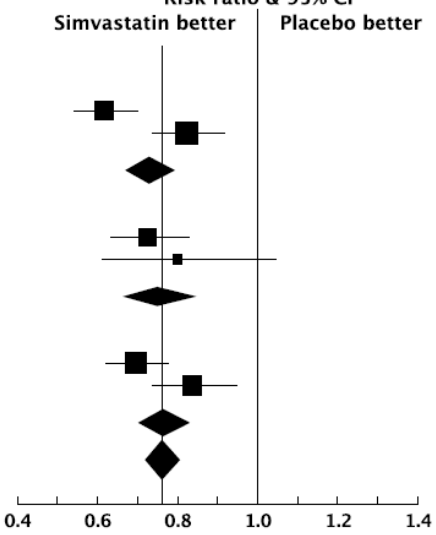
	Active (10,000)	**Control** (10,000)	**OR** (& 95%CI)	**Z score**
True events	800	1000	0.78 (0.71-0.86)	4.9
Extra false events (evenly distributed)				
+ 10%	890	1090	0.80 (0.73-0.88)	4.7
+ 20%	980	1180	0.81 (0.74-0.89)	4.6
Missing real events (evenly distributed)				
- 10%	720	900	0.78 (0.71-0.87)	4.7
- 20%	640	800	0.79 (0.71-0.88)	4.4

Effect of adjudication on assessment of efficacy:

Effect of simvastatin on major vascular events

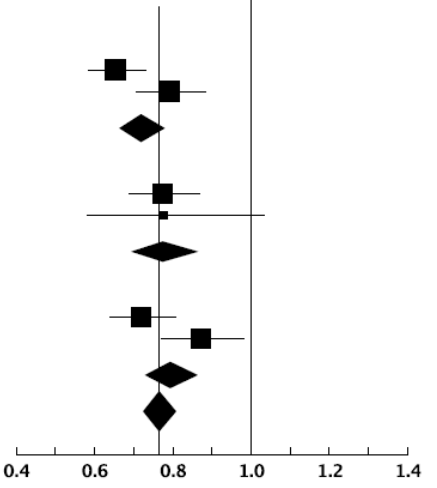
Adjudicated events

Type of event	Simvastatin allocation: Simvastatin (n=10269)		Placebo (n=10267)		Risk ratio & 95% CI
Coranary events					
Non-fatal MI	357	(3.5%)	574	(5.6%)	0.62 (0.54-0.70)
Coronary death	587	(5.7%)	707	(6.9%)	0.82 (0.74-0.92)
Any coronary event	**898**	**(8.7%)**	**1212**	**(11.8%)**	**0.73 (0.67-0.79)**
Strokes					
Non-fatal stroke	366	(3.6%)	499	(4.9%)	0.72 (0.63-0.83)
Fatal stroke	96	(0.9%)	119	(1.2%)	0.80 (0.61-1.05)
Any stroke	**444**	**(4.3%)**	**585**	**(5.7%)**	**0.75 (0.66-0.85)**
Revascularisations					
Coronary	513	(5.0%)	725	(7.1%)	0.70 (0.62-0.78)
Non-coronary	450	(4.4%)	532	(5.2%)	0.84 (0.74-0.95)
Any revascularisation	**939**	**(9.1%)**	**1205**	**(11.7%)**	**0.76 (0.70-0.83)**
ANY MAJOR VASCULAR EVENT	**2033**	**(19.8%)**	**2585**	**(25.2%)**	**0.76 (0.72-0.81)**



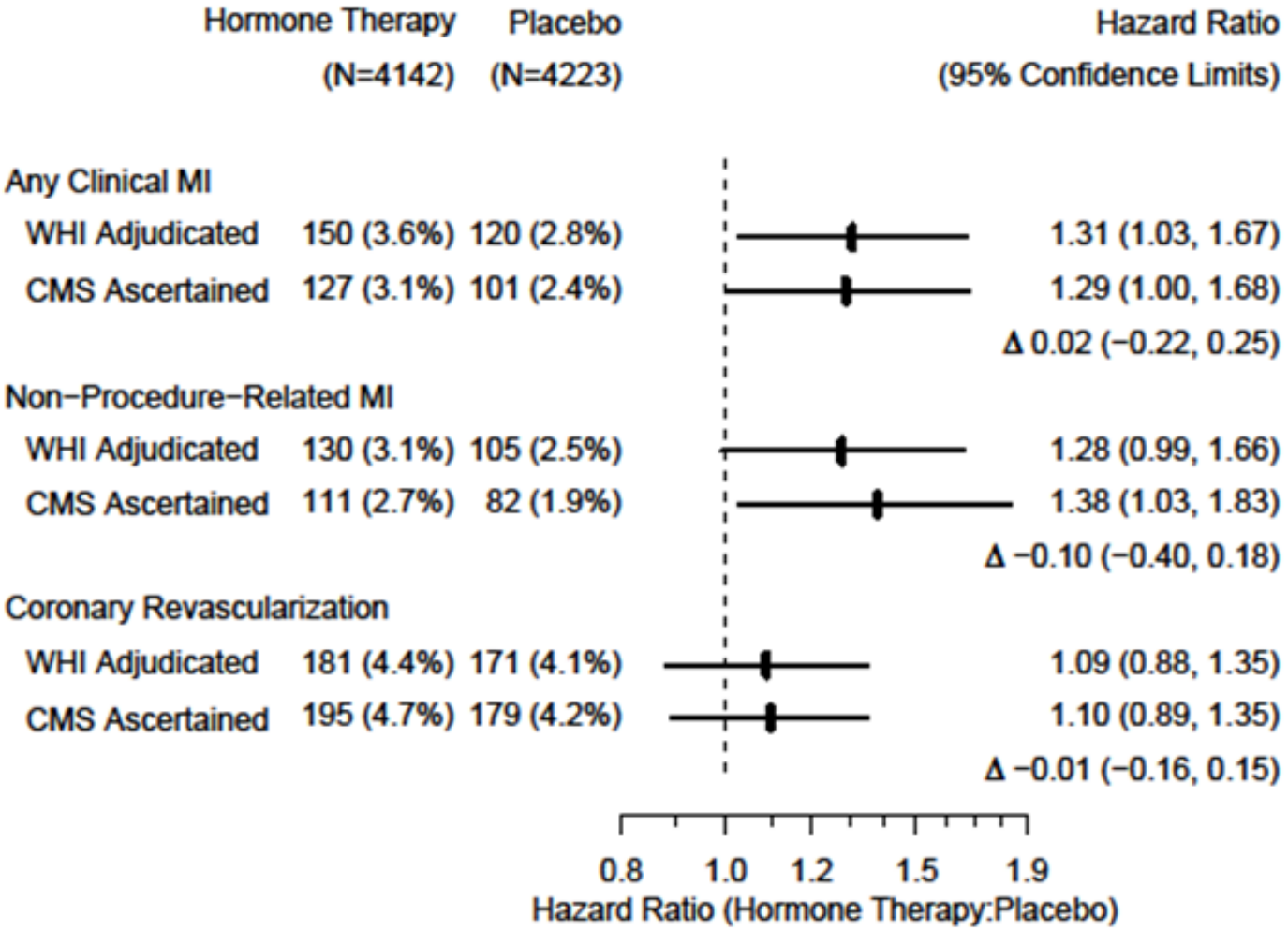
Unadjudicated events

Type of event	Simvastatin allocation: Simvastatin (n=10269)		Placebo (n=10267)		Risk ratio & 95% CI
Coranary events					
Non-fatal MI	492	(4.8%)	743	(7.2%)	0.65 (0.58-0.73)
Coronary death	547	(5.3%)	687	(6.7%)	0.79 (0.71-0.88)
Any coronary event	**988**	**(9.6%)**	**1350**	**(13.1%)**	**0.72 (0.66-0.78)**
Strokes					
Non-fatal stroke	487	(4.7%)	621	(6.0%)	0.77 (0.69-0.87)
Fatal stroke	82	(0.8%)	105	(1.0%)	0.78 (0.58-1.03)
Any stroke	**550**	**(5.4%)**	**700**	**(6.8%)**	**0.77 (0.69-0.87)**
Revascularisations					
Coronary	487	(4.7%)	667	(6.5%)	0.72 (0.64-0.81)
Non-coronary	492	(4.8%)	559	(5.4%)	0.87 (0.77-0.98)
Any revascularisation	**943**	**(9.2%)**	**1166**	**(11.4%)**	**0.79 (0.73-0.86)**
ANY MAJOR VASCULAR EVENT	**2187**	**(21.3%)**	**2765**	**(26.9%)**	**0.77 (0.72-0.81)**



Using adjudicated vs routine claims data:

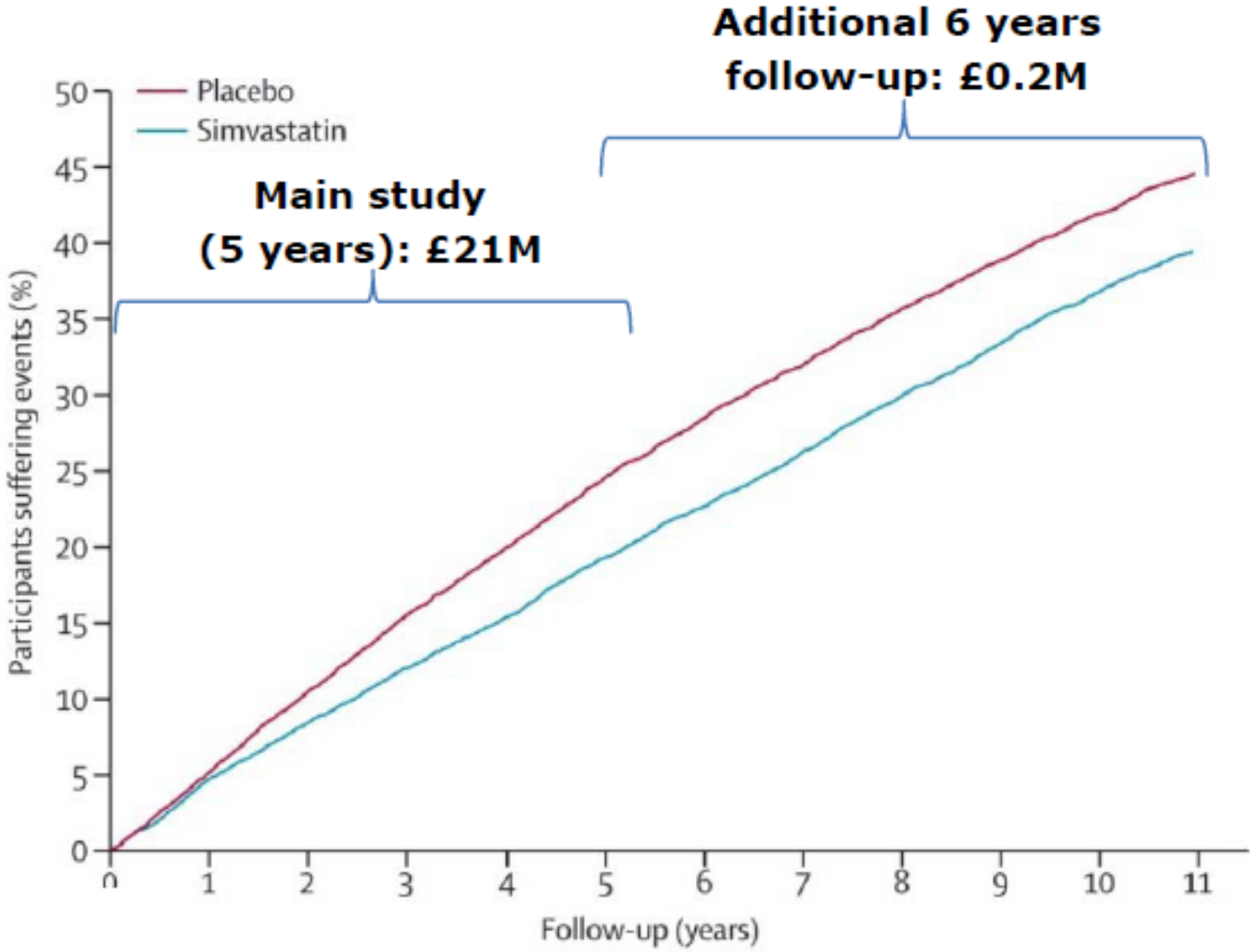
Effect of HRT on cardiac events in Women's Health Initiative



Hlatky et al. Circ Cardiovasc Qual Outcomes 2014

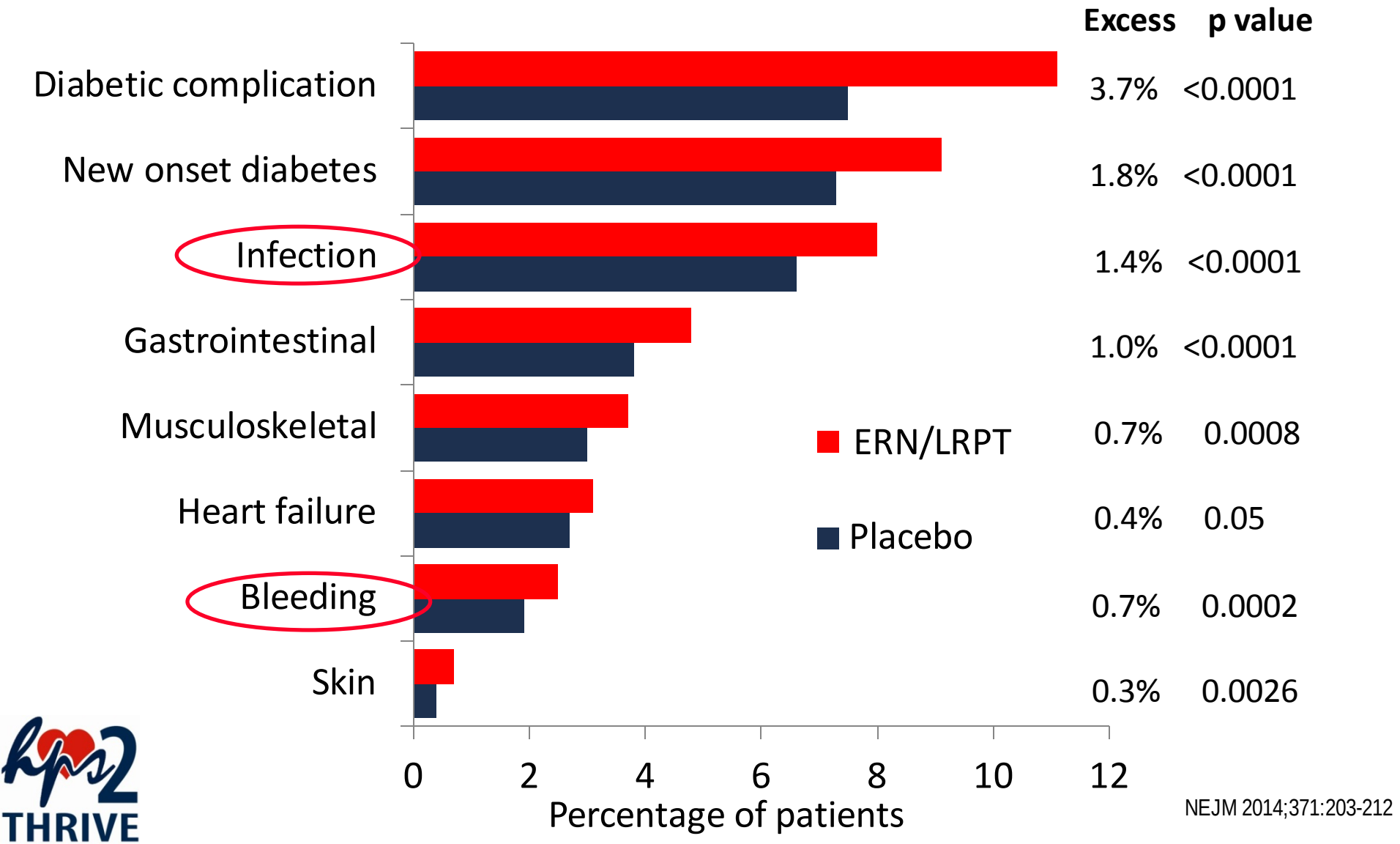
Routine data to assess long-term treatment effects:
Lowering cholesterol reduces risk of vascular events

20,536 patients randomized to simvastatin vs placebo in the Heart Protection Study



Heart Protection Study Collaborative Group Lancet 2011

Adverse effects of niacin/laropiprant identified from self-reported, unadjudicated serious adverse events



THRIVE

NEJM 2014;371:203-212

New Challenges

- **Information governance**
  - de-identification, anonymization & record linkage
  - data protection vs data transparency
  - Key: build trust - clear purpose & defined controls
- **Technical**
  - data acquisition, cleaning, visualization & analysis
  - interoperability, data standards, legacy systems, etc
  - technical failures (e.g. connectivity, GPS, battery issues)
- **Methodological**
  - clarification & categorization of endpoints
  - role of adjudication?

Implications for Regulatory Science

- **Interpretation and decision-making**
  - what is the impact of missing data – bias vs noise?
    (e.g. technical failure, app-fatigue, migration)
  - what is a valid endpoint?
  - how to validate novel endpoints?
  - regulatory & clinical decisions are typically binary

- **Good Clinical Practice**
  - Trials without borders / trials in the wild
  - Outdated & unsuitable detailed GCP requirements
  - Source data: what is it? will it change? who controls it? archive?

GCP for 21st Century Clinical Trials

- **Facilitate efficient, high quality assessment of health interventions**
- **Focus on the key principles** (not operational details), i.e. the avoidance of errors that matter to decision making regarding:
  - the rights, safety & wellbeing of participants
  - the reliability of the results
  - wider public health & the environment
- **Embrace & encourage methodological & technological innovation**
- **Proportionate, efficient and coordinated** with other research and clinical governance requirements
- **Developed, reviewed & revised by all stakeholders** including participants, trialists, regulators, industry

Conclusion

- **High quality randomized clinical trials are essential for regulatory & clinical decision making**
- **Fundamental principles of randomized trials remain**
- **Big Data offers potential to assess effects at scale, breadth & duration:**
  - Efficient methods to assess existing outcome concepts
  - New methods to assess novel outcome concept
- **Significant challenges:**
  - Technical / methodological
  - Information governance
  - Regulatory decision making
  - Outdated GCP and other trial conduct regulations