Highlights from the CM case study data

COST TRAINING SCHOOL

WARSAW

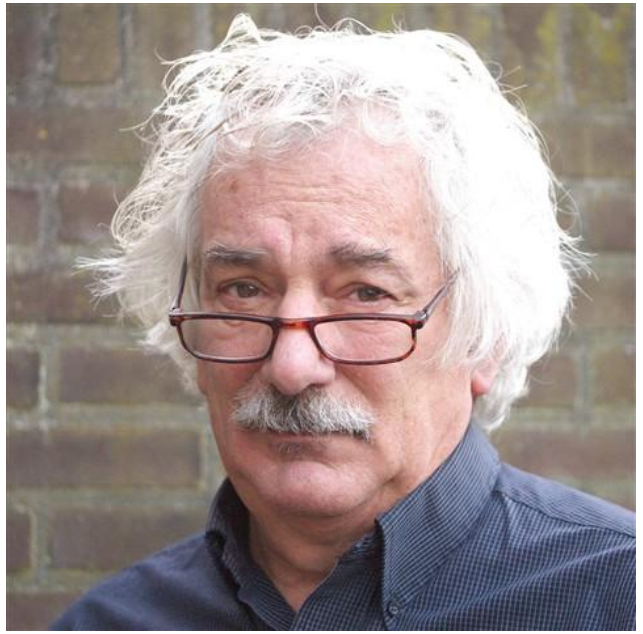
MARCH 16-18, 2016

ABIGAIL COLSON

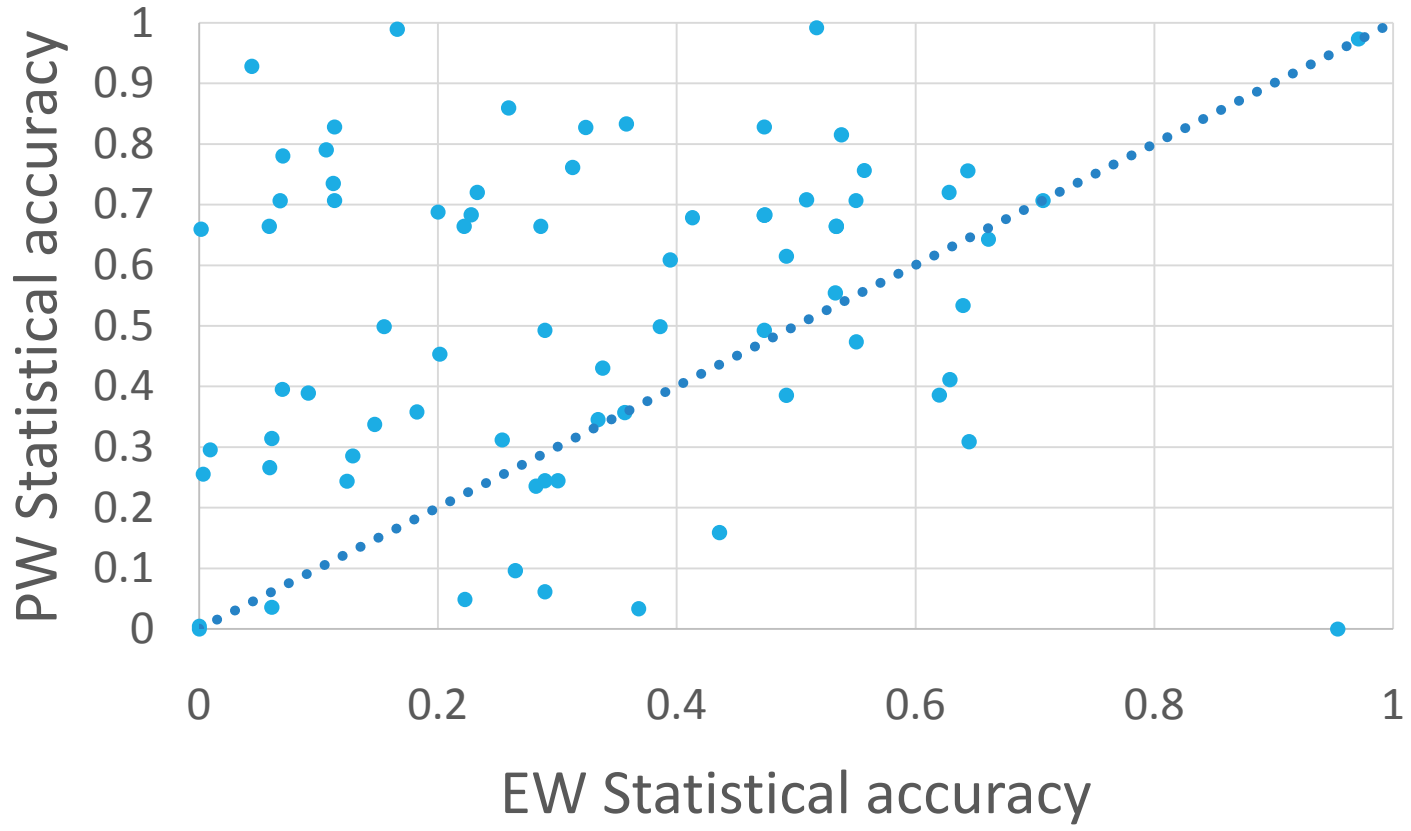
UNIVERSITY OF STRATHCLYDE

Preface: Some useful resources

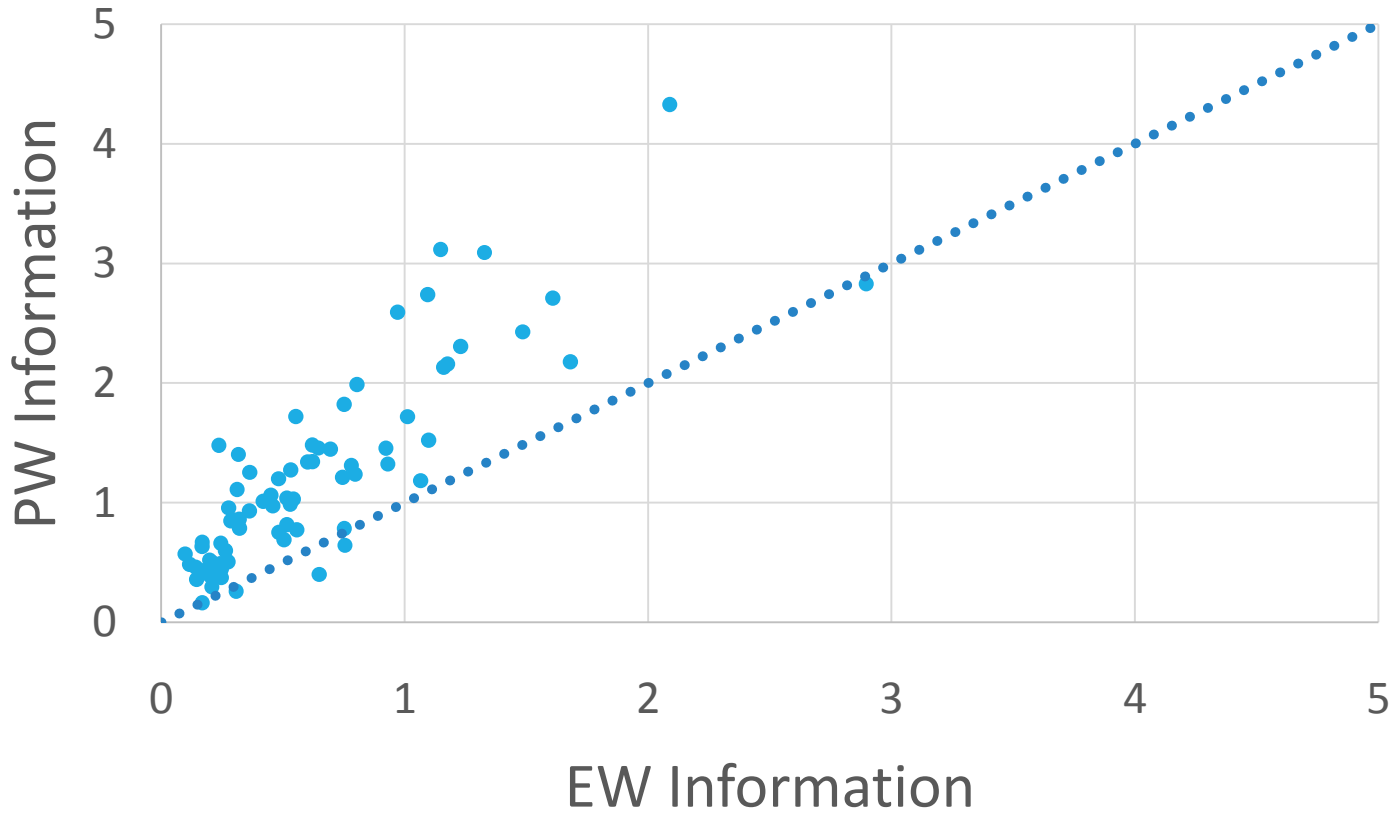
- Cooke, Roger M., and Louis L. H. J. Goossens. 2008. "TU Delft Expert Judgment Data Base." *Reliability Engineering & System Safety*, Expert Judgement, 93 (5): 657–74. doi:10.1016/j.ress.2007.03.005.
- http://rogermcooke.net/
  - Excalibur files for everything in the above paper!
  - Also Excalibur files for 33 post-2006 studies!
  - Some relevant recent papers and working papers!
  - An audio file of Roger playing bass.
  - Miscellaneous other things.



Why use performance weights?



Why use performance weights?



Can you get a high-scoring performance-weight decision maker from a group of low-scoring experts?

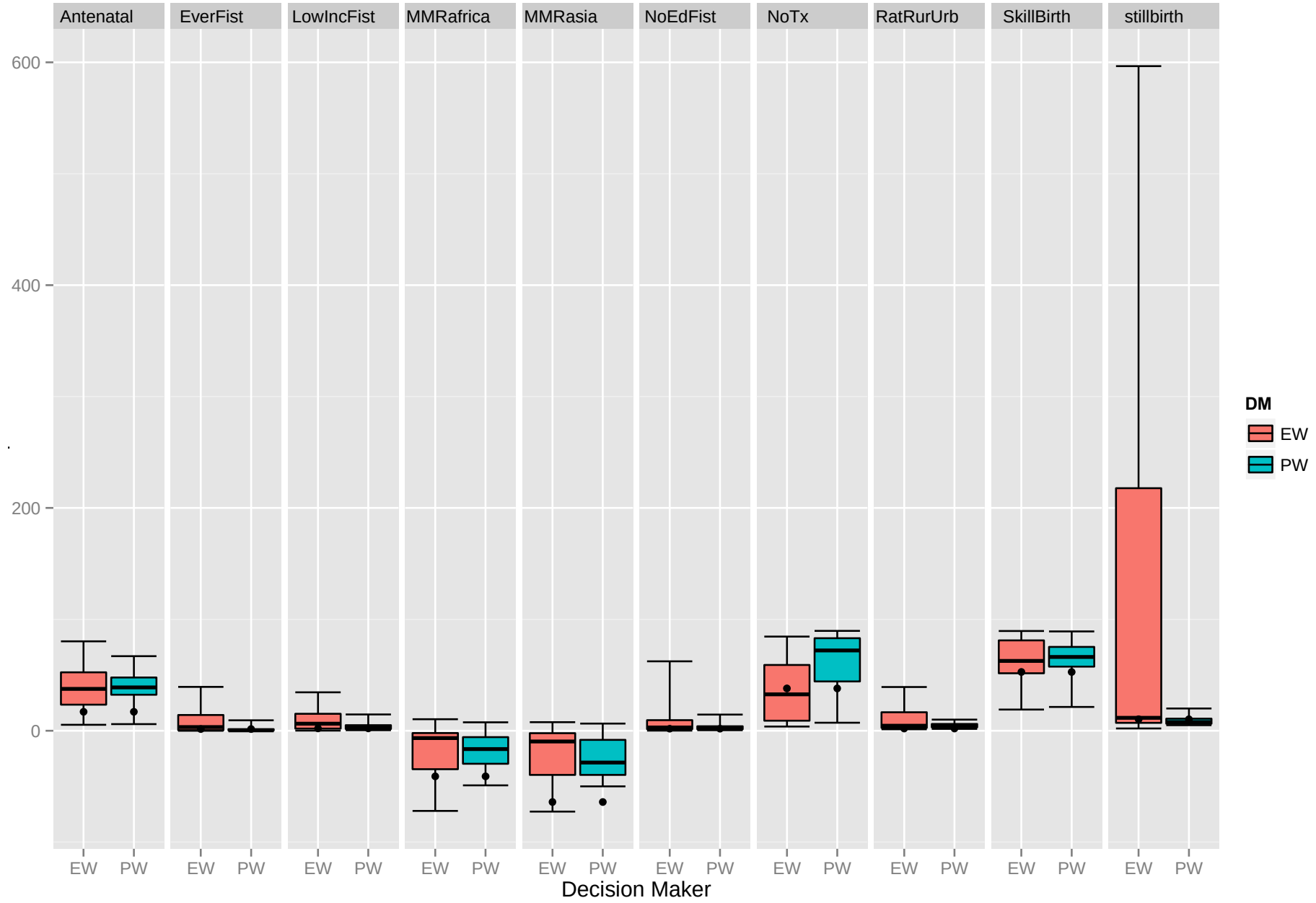
Can you get a high-scoring performance-weight decision maker from a group of low-scoring experts?

YES!

Obstetric fistula

Expert	P-value	Information	Weight
Expert1	0.0002059	1.331	0
Expert2	0.0001328	2.017	0
Expert3	6.23E-06	3.126	0
Expert4	0.005928	2.662	0.357
Expert5	0.007621	1.267	0.597
Expert6	2.40E-05	1.558	0
Expert7	0.0005007	4.181	0.0467
Expert8	5.25E-05	2.658	0
PW	0.2659	1.776	-
EW	0.05891	0.7006	-

Colson, Abigail R., Sweta Adhikari, Ambereen Sleemi, and Ramanan Laxminarayan. 2015. “Quantifying Uncertainty in Intervention Effectiveness with Structured Expert Judgement: An Application to Obstetric Fistula.” *BMJ Open* 5 (6): e007233. doi:10.1136/bmjopen-2014-007233.


Antenatal
EverFist
LowIncFist
MMRafrica
MMRasia
NoEdFist
NoTx
RatRurUrb
SkillBirth
stillbirth
600
400
200
0
EW
PW
Decision Maker
DM
EW
PW

Does the “best expert” always get weight in the optimized decision maker?

Does the “best expert” always get weight in the optimized decision maker?

NO. THIS HAPPENED IN 1 POST-2006 CASE.

(IT CAN ALSO HAPPEN THAT THE BEST EXPERT OUTPERFORMS THE OPTIMIZED DECISION MAKER.)

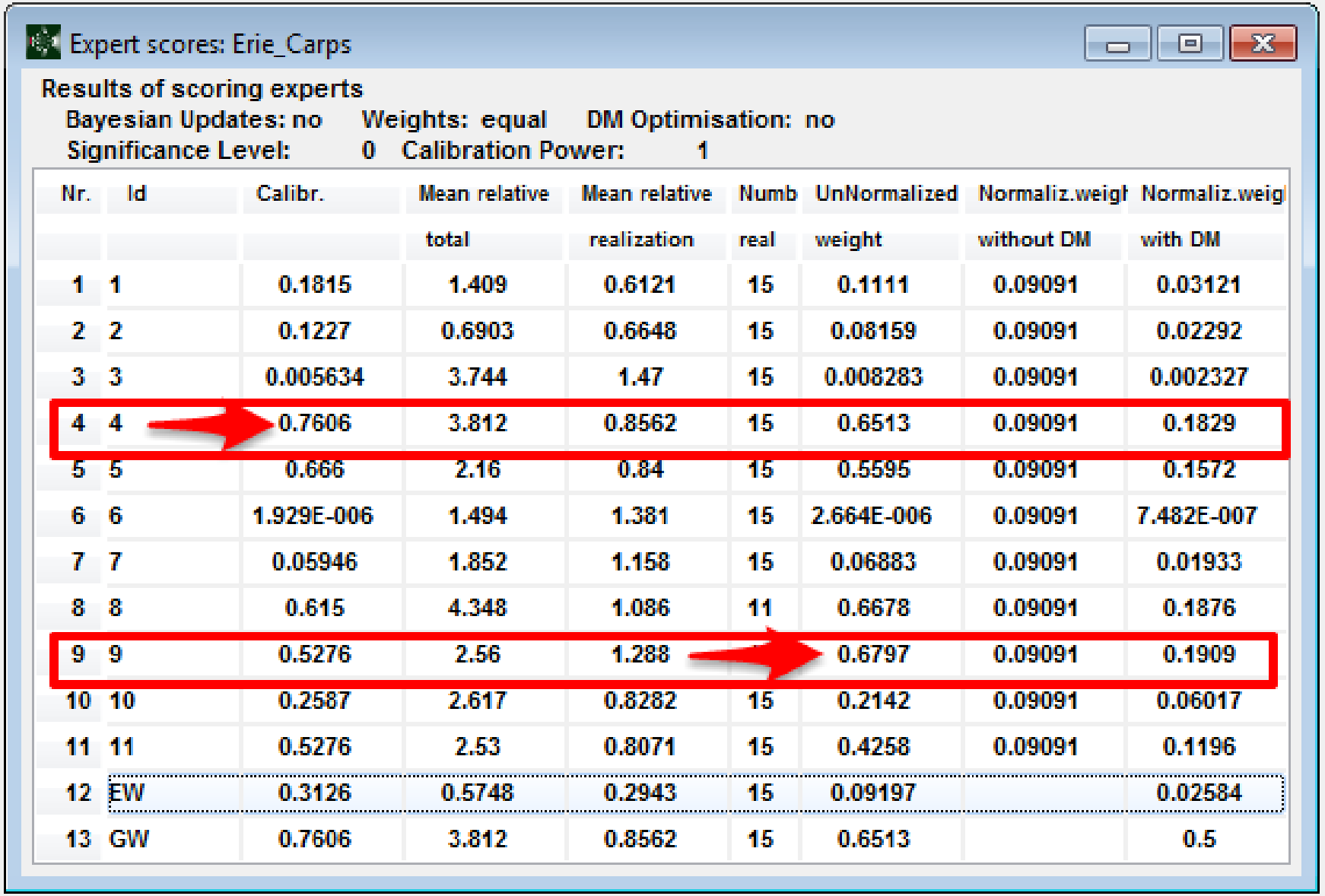
Expert scores: Erie_Carps

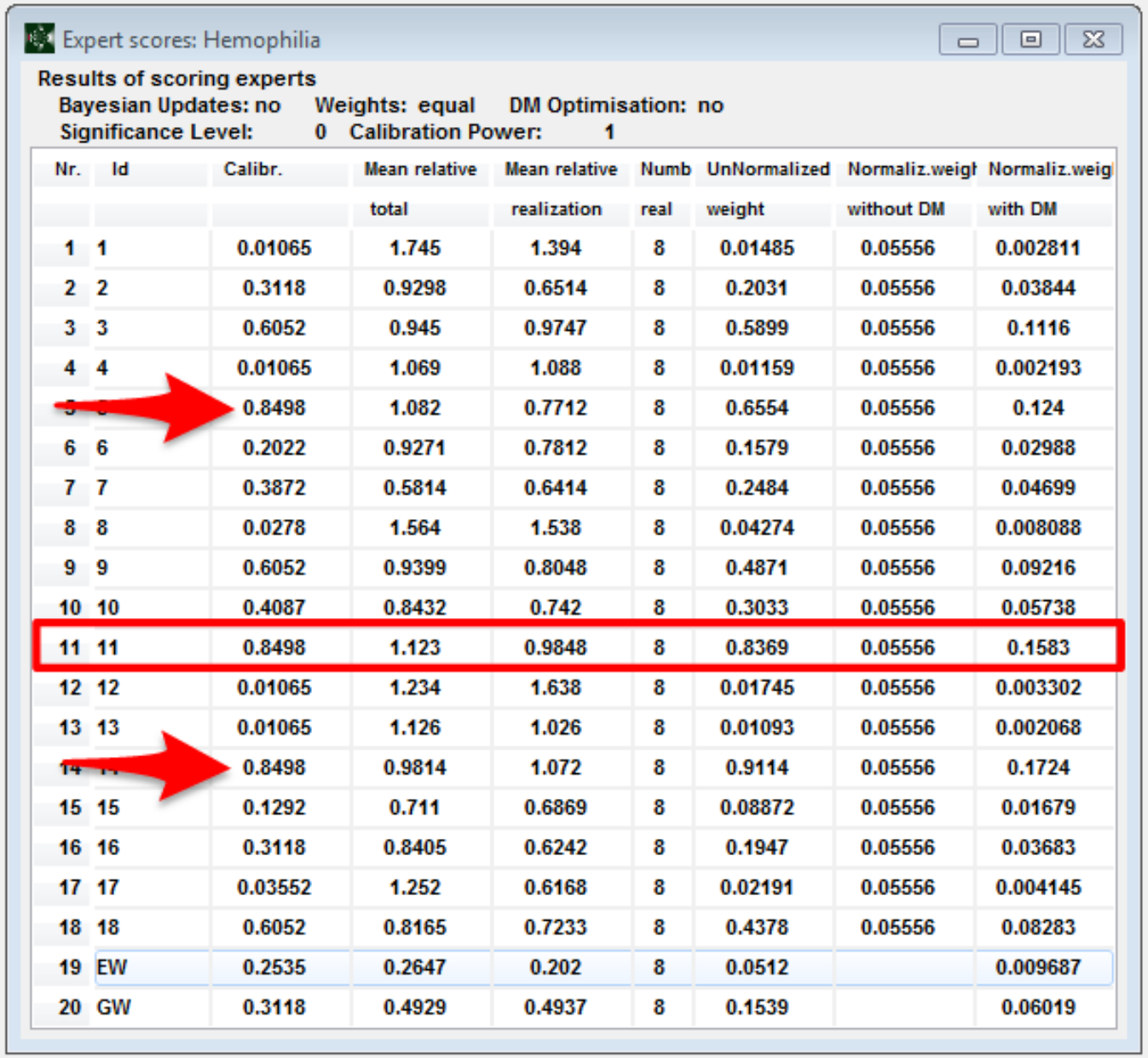
Results of scoring experts

Bayesian Updates: no Weights: global DM Optimisation: yes

Significance Level: 0.7606 Calibration Power: 1

Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weigt without DM	Normaliz.weig with DM
1	1	0.1815	1.409	0.6121	15	0	0	0
2	2	0.1227	0.6903	0.6648	15	0	0	0
3	3	0.005634	3.744	1.47	15	0	0	0
4	4	0.7606	3.812	0.8562	15	0.6513	1	0.5
5	5	0.666	2.16	0.84	15	0	0	0
6	6	1.929E-006	1.494	1.381	15	0	0	0
7	7	0.05946	1.852	1.158	15	0	0	0
8	8	0.615	4.348	1.086	11	0	0	0
9	9	0.5276	2.56	1.288	15	0	0	0
10	10	0.2587	2.617	0.8282	15	0	0	0
11	11	0.5276	2.53	0.8071	15	0	0	0
12	GW	0.7606	3.812	0.8562	15	0.6513		0.5

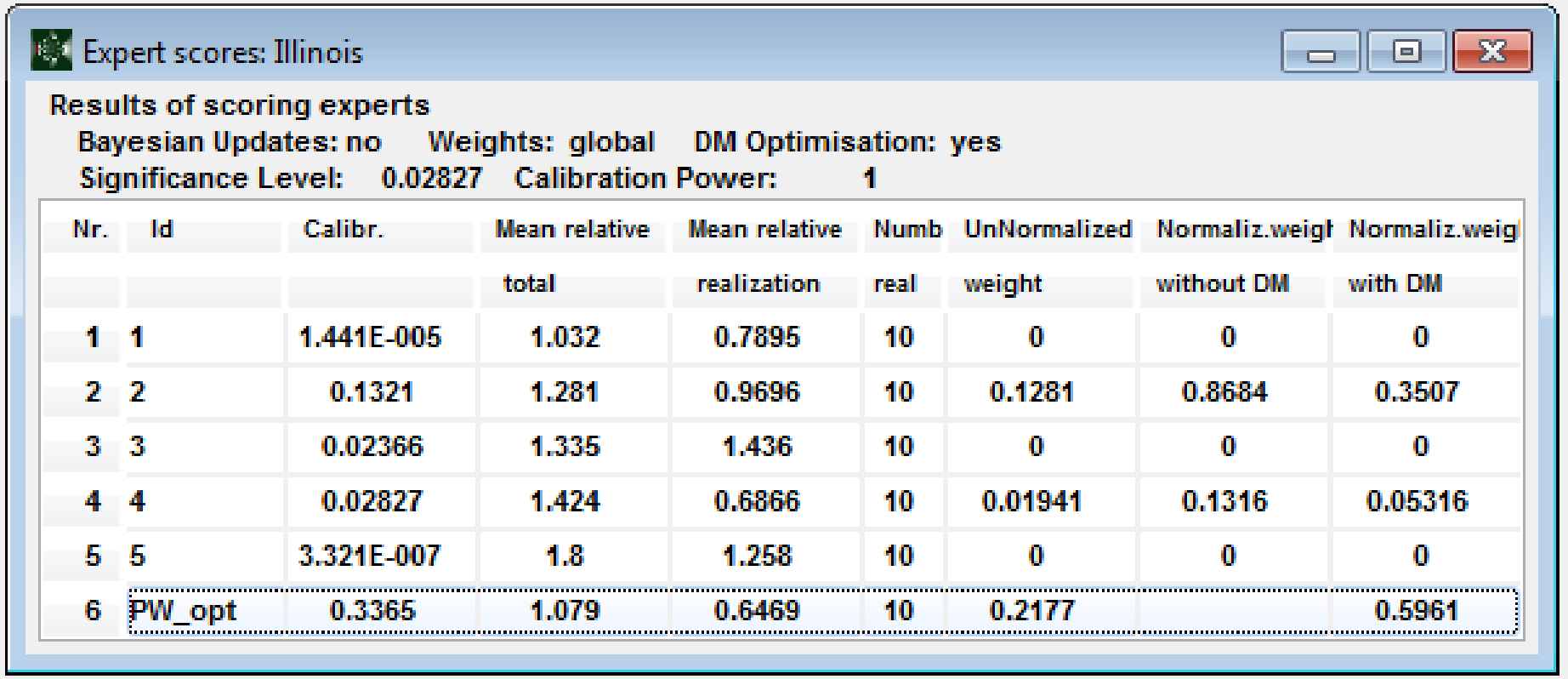

Expert scores: Erie_Carps
Results of scoring experts
Bayesian Updates: no Weights: equal DM Optimisation: no
Significance Level: 0 Calibration Power: 1
Nr. Id Calibr. Mean relative total Mean relative realization Numb real UnNormalized weight Normaliz.weigh without DM Normaliz.weig with DM
1 1 0.1815 1.409 0.6121 15 0.1111 0.09091 0.03121
2 2 0.1227 0.6903 0.6648 15 0.08159 0.09091 0.02292
3 3 0.005634 3.744 1.47 15 0.008283 0.09091 0.002327
4 4 0.7606 3.812 0.8562 15 0.6513 0.09091 0.1829
5 5 0.666 2.16 0.84 15 0.5595 0.09091 0.1572
6 6 1.929E-006 1.494 1.381 15 2.664E-006 0.09091 7.482E-007
7 7 0.05946 1.852 1.158 15 0.06883 0.09091 0.01933
8 8 0.615 4.348 1.086 11 0.6678 0.09091 0.1876
9 9 0.5276 2.56 1.288 0.6797 0.09091 0.1909
10 10 0.2587 2.617 0.8282 15 0.2142 0.09091 0.06017
11 11 0.5276 2.53 0.8071 15 0.4258 0.09091 0.1196
12 EW 0.3126 0.5748 0.2943 15 0.09197 0.02584
13 GW 0.7606 3.812 0.8562 15 0.6513 0.5

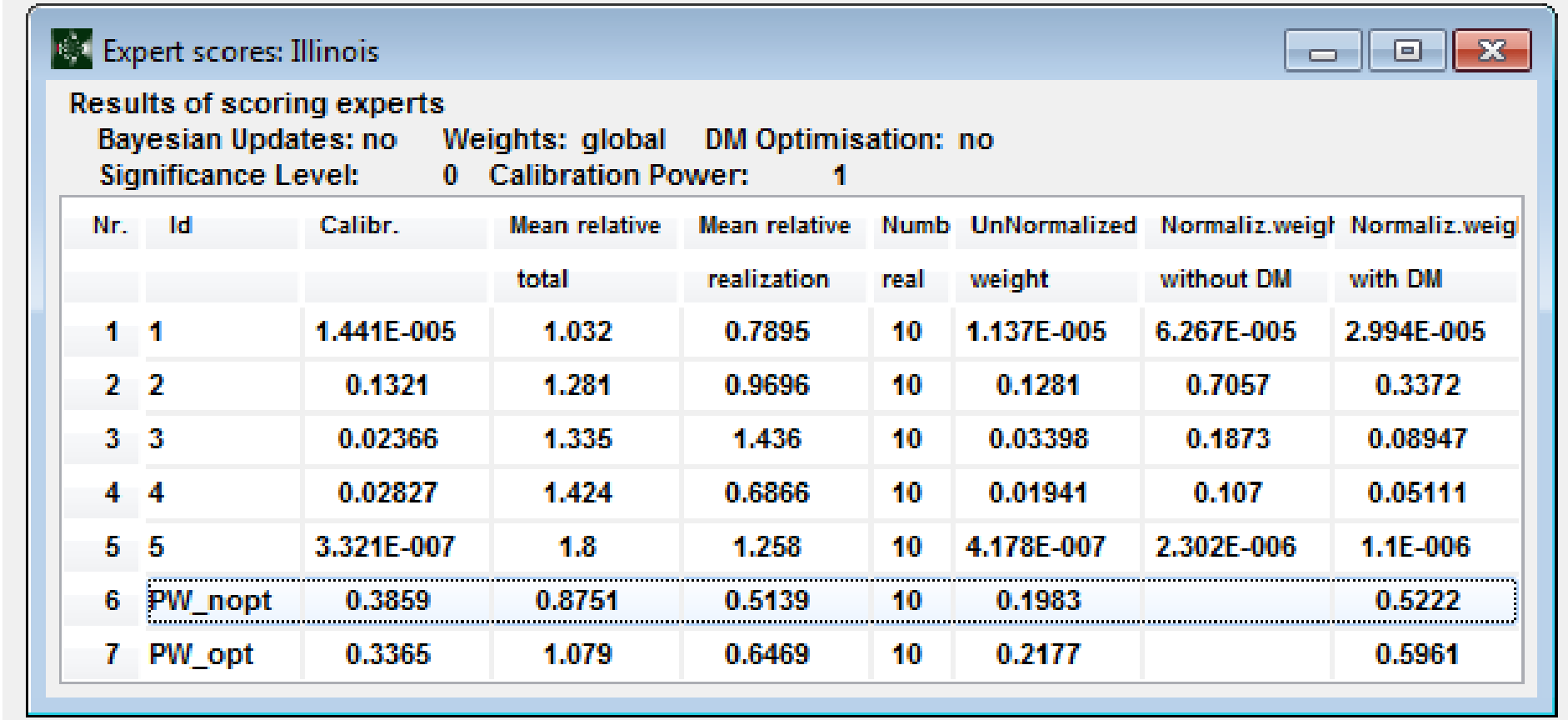

Expert scores: Hemophilia
Results of scoring experts
Bayesian Updates: no Weights: equal DM Optimisation: no
Significance Level: 0 Calibration Power: 1
Nr. Id Calibr. Mean relative total Mean relative realization Numb real UnNormalized weight Normaliz.weigh without DM Normaliz.weig with DM
1 1 0.01065 1.745 1.394 8 0.01485 0.05556 0.002811
2 2 0.3118 0.9298 0.6514 8 0.2031 0.05556 0.03844
3 3 0.6052 0.945 0.9747 8 0.5899 0.05556 0.1116
4 4 0.01065 1.069 1.088 8 0.01159 0.05556 0.002193
0.8498 1.082 0.7712 8 0.6554 0.05556 0.124
6 6 0.2022 0.9271 0.7812 8 0.1579 0.05556 0.02988
7 7 0.3872 0.5814 0.6414 8 0.2484 0.05556 0.04699
8 8 0.0278 1.564 1.538 8 0.04274 0.05556 0.008088
9 9 0.6052 0.9399 0.8048 8 0.4871 0.05556 0.09216
10 10 0.4087 0.8432 0.742 8 0.3033 0.05556 0.05738
11 11 0.8498 1.123 0.9848 8 0.8369 0.05556 0.1583
12 12 0.01065 1.234 1.638 8 0.01745 0.05556 0.003302
13 13 0.01065 1.126 1.026 8 0.01093 0.05556 0.002068
0.8498 0.9814 1.072 8 0.9114 0.05556 0.1724
15 15 0.1292 0.711 0.6869 8 0.08872 0.05556 0.01679
16 16 0.3118 0.8405 0.6242 8 0.1947 0.05556 0.03683
17 17 0.03552 1.252 0.6168 8 0.02191 0.05556 0.004145
18 18 0.6052 0.8165 0.7233 8 0.4378 0.05556 0.08283
19 EW 0.2535 0.2647 0.202 8 0.0512 0.009687
20 GW 0.3118 0.4929 0.4937 8 0.1539 0.06019

What’s the difference between optimized and non-optimized performance-weights?

What’s the difference between optimized and non-optimized performance-weights?

THERE **CAN BE** A BIG PERFORMANCE DIFFERENCE BETWEEN THE TWO OPTIONS.


Expert scores: Illinois
Results of scoring experts
Bayesian Updates: no Weights: global DM Optimisation: yes
Significance Level: 0.02827 Calibration Power: 1
Nr. Id Calibr. Mean relative total Mean relative realization Numb real UnNormalized weight Normaliz.weigh without DM Normaliz.weig with DM
1 1 1.441E-005 1.032 0.7895 10 0 0 0
2 2 0.1321 1.281 0.9696 10 0.1281 0.8684 0.3507
3 3 0.02366 1.335 1.436 10 0 0 0
4 4 0.02827 1.424 0.6866 10 0.01941 0.1316 0.05316
5 5 3.321E-007 1.8 1.258 10 0 0 0
6 PW_opt 0.3365 1.079 0.6469 10 0.2177 0.5961



Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weigh without DM	Normaliz.weig with DM
1	1	1.441E-005	1.032	0.7895	10	1.137E-005	6.267E-005	2.994E-005
2	2	0.1321	1.281	0.9696	10	0.1281	0.7057	0.3372
3	3	0.02366	1.335	1.436	10	0.03398	0.1873	0.08947
4	4	0.02827	1.424	0.6866	10	0.01941	0.107	0.05111
5	5	3.321E-007	1.8	1.258	10	4.178E-007	2.302E-006	1.1E-006
6	PW_nopt	0.3859	0.8751	0.5139	10	0.1983		0.5222
7	PW_opt	0.3365	1.079	0.6469	10	0.2177		0.5961

Expert scores: CDC_ROI

Results of scoring experts

Bayesian Updates: no Weights: global DM Optimisation: yes

Significance Level: 0.7203 Calibration Power: 1

Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weigh without DM	Normaliz.weig with DM
1	1	0.7203	2.597	2.305	10	1.66	1	0.5
2	2	1.602E-005	1.904	1.655	10	0	0	0
3	3	1.273E-006	2.344	3.49	10	0	0	0
4	4	5.559E-006	2.961	2.719	10	0	0	0
5	5	0.4988	2.341	1.39	10	0	0	0
6	6	0.01651	1.39	1.355	10	0	0	0
7	7	2.181E-007	2.09	3.345	10	0	0	0
8	8	0.4988	3.825	1.737	10	0	0	0
9	9	0.1321	4.623	1.719	10	0	0	0
10	10	1.273E-006	3.08	3.071	10	0	0	0
11	11	0.02366	4.113	1.82	10	0	0	0
12	12	0.00917	2.797	2.304	10	0	0	0
13	13	0.007147	2.758	2.063	10	0	0	0
14	14	0.0001328	3.815	3.279	10	0	0	0
15	15	0.1249	3.843	2.66	10	0	0	0
16	16	0.0003053	2.79	2.057	10	0	0	0
17	17	0.02919	2.471	1.745	10	0	0	0
18	18	1.428E-006	2.727	3.155	10	0	0	0
19	19	0.4988	2.032	1.91	10	0	0	0
20	20	0.04675	2.208	2.183	10	0	0	0
21	PW_opt	0.7203	2.597	2.305	10	1.66		0.5

Expert scores: CDC_ROI

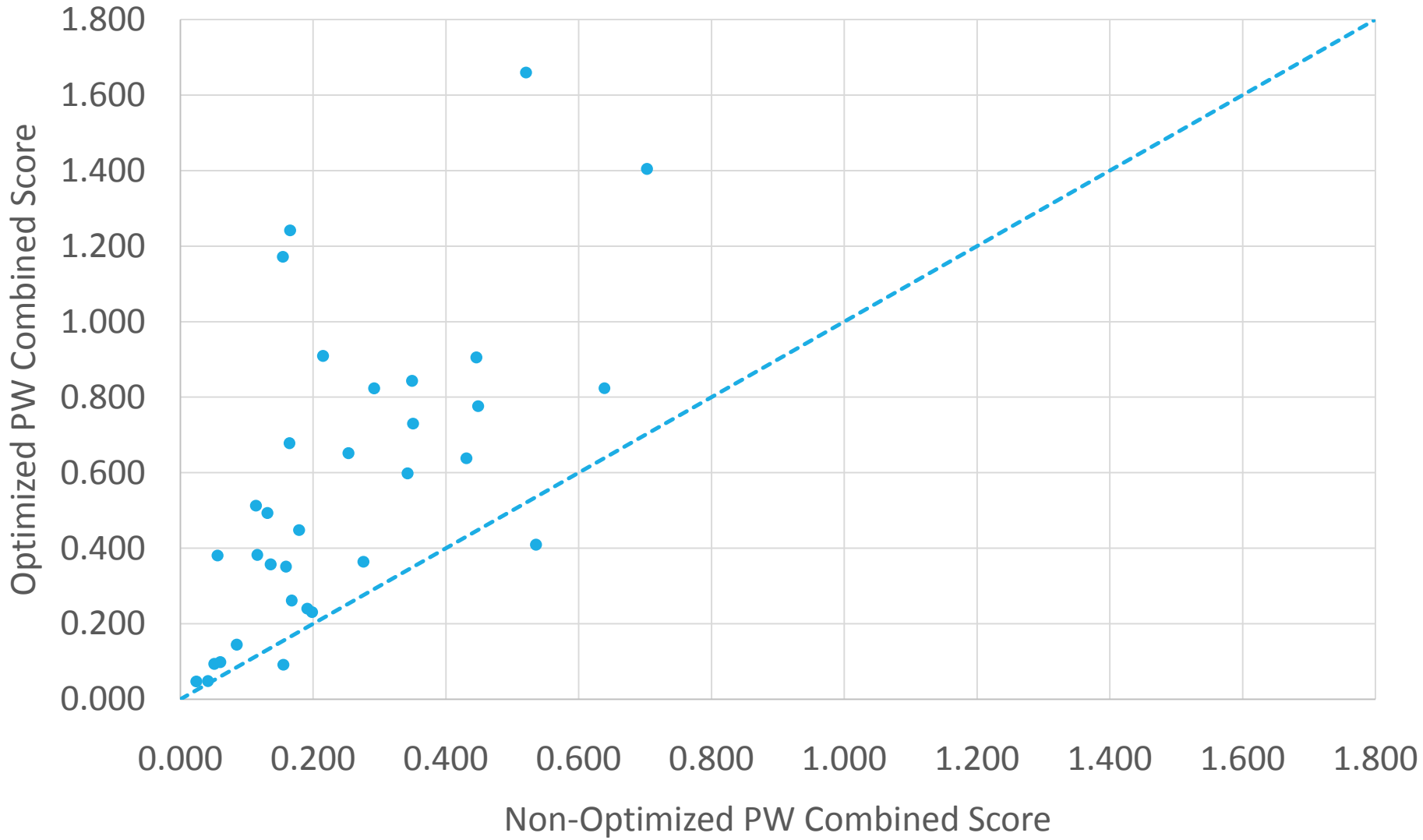
Results of scoring experts

Bayesian Updates: no Weights: global DM Optimisation: no

Significance Level: 0 Calibration Power: 1

Nr.	Id	Calibr.	Mean relative total	Mean relative realization	Numb real	UnNormalized weight	Normaliz.weigh without DM	Normaliz.weig with DM
1	1	0.7203	2.597	2.305	10	1.66	0.3329	0.3014
2	2	1.602E-005	1.904	1.655	10	2.651E-005	5.315E-006	4.812E-006
3	3	1.273E-006	2.344	3.49	10	4.443E-006	8.907E-007	8.065E-007
4	4	5.559E-006	2.961	2.719	10	1.511E-005	3.03E-006	2.744E-006
5	5	0.4988	2.341	1.39	10	0.6936	0.1391	0.1259
6	6	0.01651	1.39	1.355	10	0.02237	0.004484	0.00406
7	7	2.181E-007	2.09	3.345	10	7.294E-007	1.462E-007	1.324E-007
8	8	0.4988	3.825	1.737	10	0.8665	0.1737	0.1573
9	9	0.1321	4.623	1.719	10	0.227	0.04552	0.04122
10	10	1.273E-006	3.08	3.071	10	3.909E-006	7.837E-007	7.096E-007
11	11	0.02366	4.113	1.82	10	0.04306	0.008634	0.007817
12	12	0.00917	2.797	2.304	10	0.02113	0.004236	0.003836
13	13	0.007147	2.758	2.063	10	0.01474	0.002956	0.002677
14	14	0.0001328	3.815	3.279	10	0.0004355	8.732E-005	7.906E-005
15	15	0.1249	3.843	2.66	10	0.3324	0.06664	0.06034
16	16	0.0003053	2.79	2.057	10	0.0006281	0.0001259	0.000114
17	17	0.02919	2.471	1.745	10	0.05092	0.01021	0.009243
18	18	1.428E-006	2.727	3.155	10	4.505E-006	9.032E-007	8.178E-007
19	19	0.4988	2.032	1.91	10	0.9526	0.191	0.1729
20	20	0.04675	2.208	2.183	10	0.1021	0.02046	0.01853
21	PW_nopt	0.3859	1.402	1.349	10	0.5208		0.09454
22	PW_opt	0.7203	2.597	2.305	10	1.66		0.5

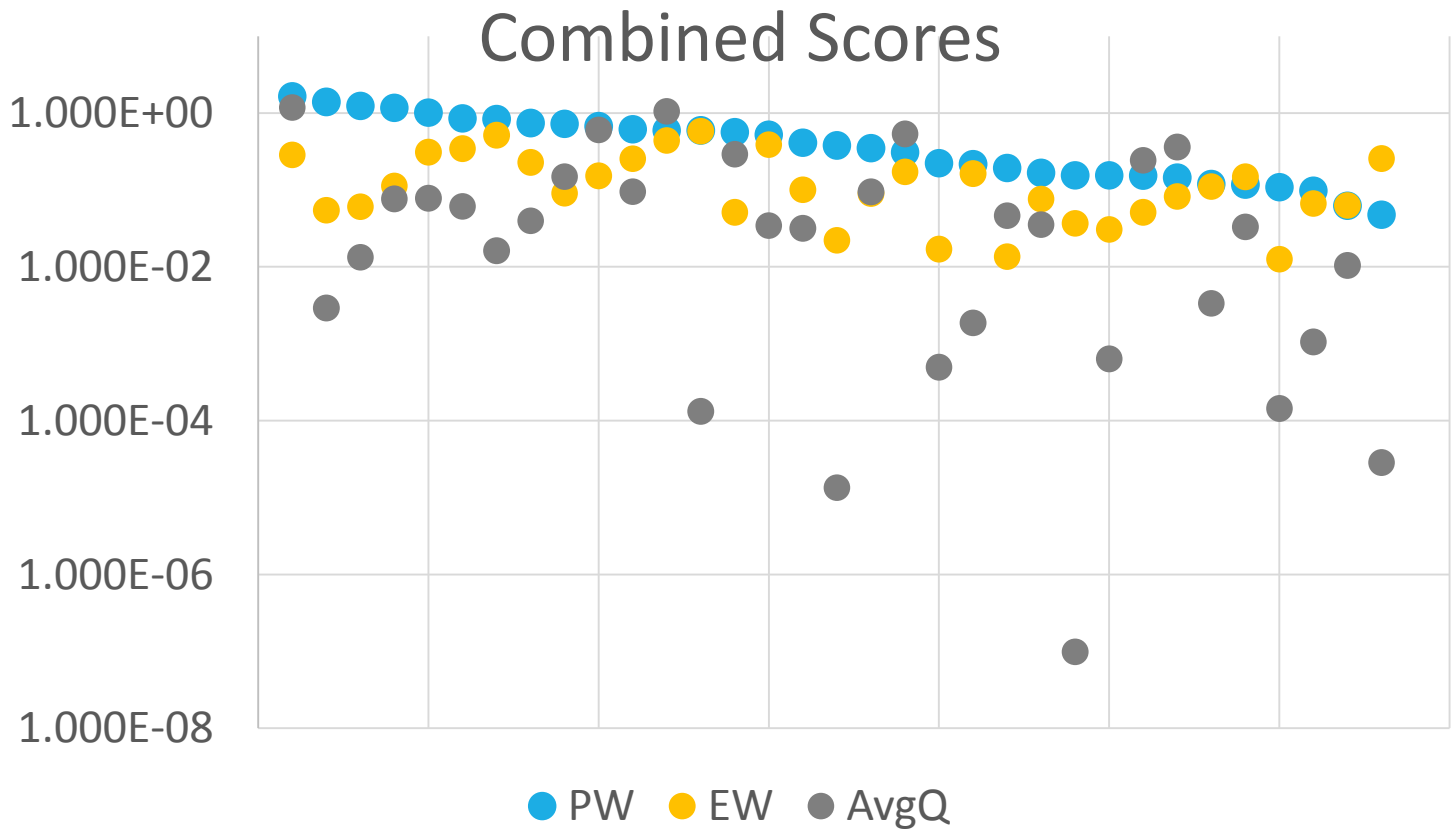
Optimizing improves the performance of the DM.



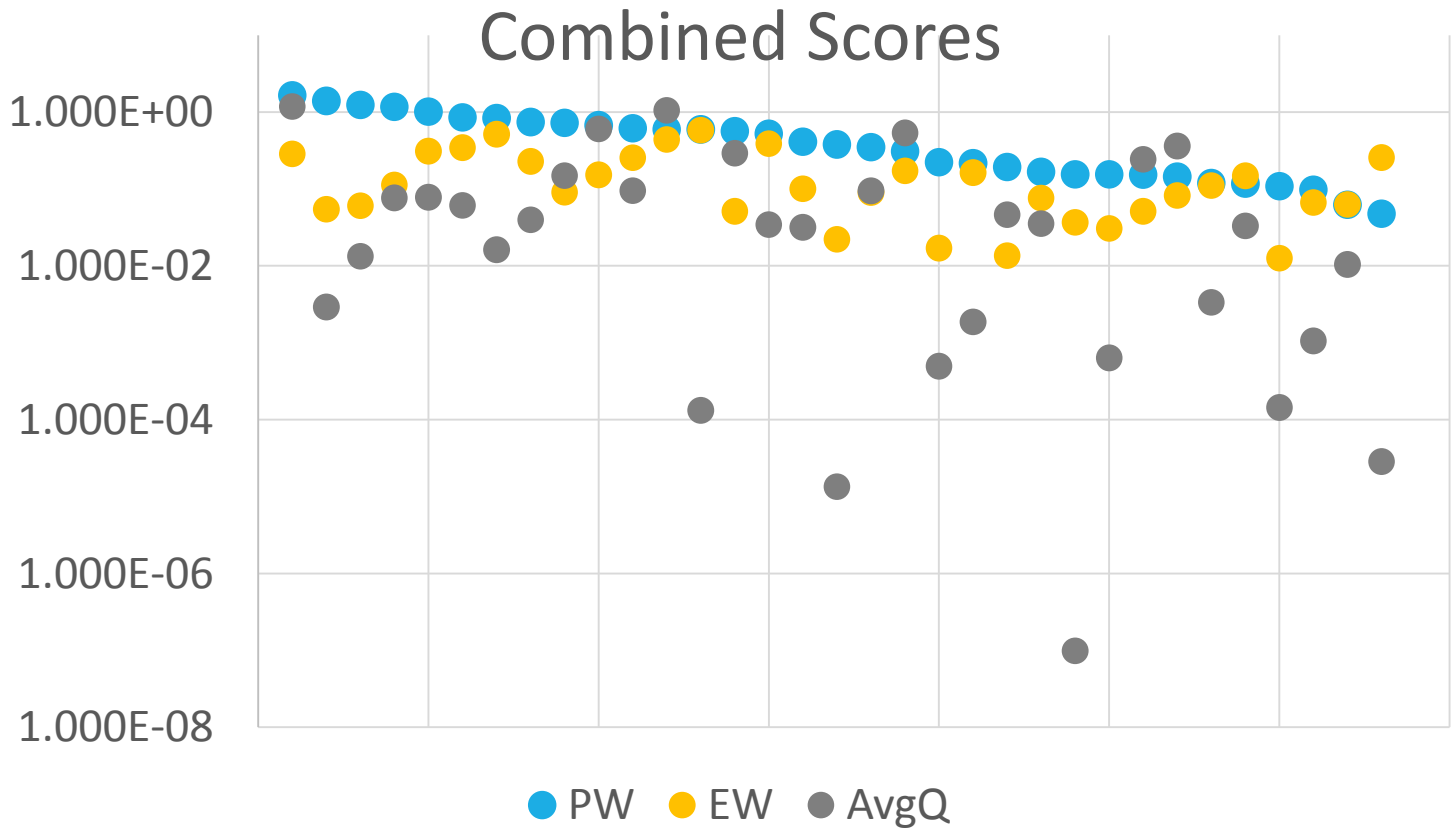
What’s the difference between averaging quantiles and averaging distributions?

Does it make a difference?

Does it make a difference?



Does it make a difference?



YES!

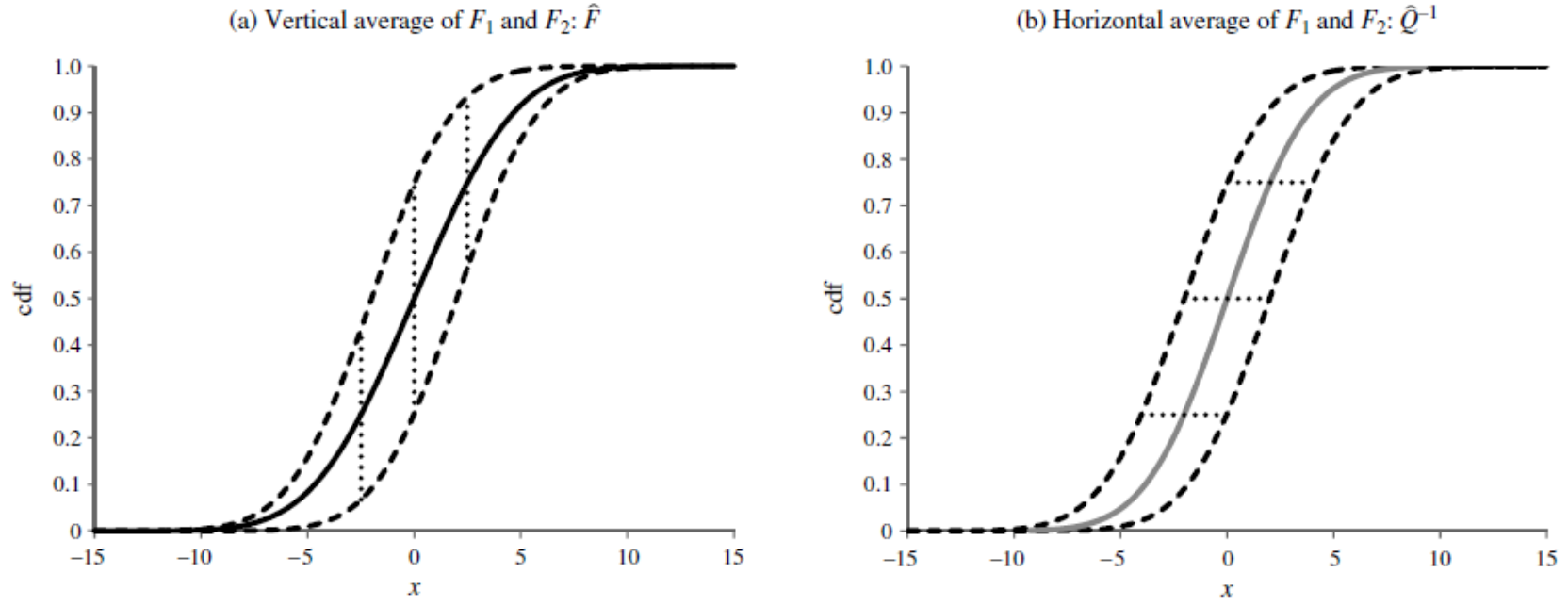
References

Lichtendahl, Kenneth C., Yael Grushka-Cockayne, and Robert L. Winkler. 2013. “Is It Better to Average Probabilities or Quantiles?” *Management Science* 59 (7): 1594–1611. doi:10.1287/mnsc.1120.1667.

Roger M. Cooke. 2015. Mathematics and Background for the Classical Model; SOM for Cross Validation. http://rogermcooke.net/rogermcooke_files/Supplementary%20Material%20for%20Cross%20Validation.pdf

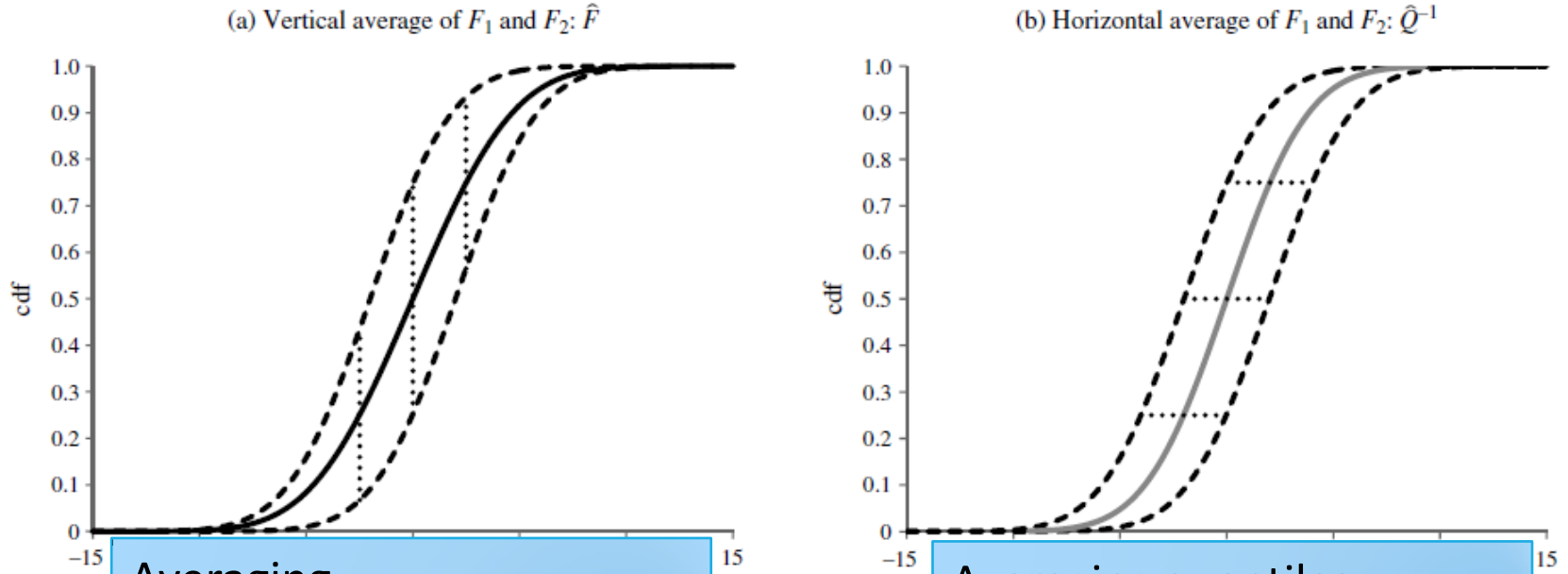
Rough Idea

Figure 2 **Comparison of $\hat{F}$ and $\hat{Q}^{-1}$ and Their Corresponding Densities from Example 1**



Rough Idea

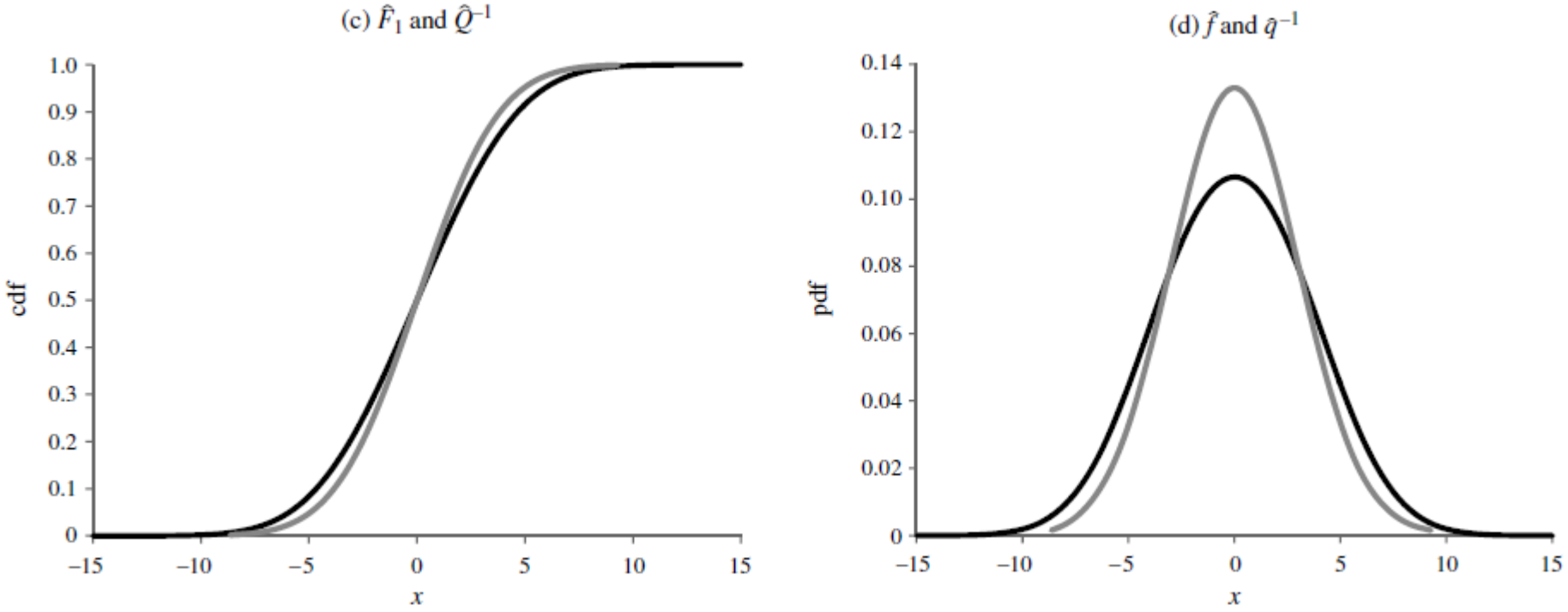
Figure 2 **Comparison of $\hat{F}$ and $\hat{Q}^{-1}$ and Their Corresponding Densities from Example 1**



Averaging probabilites/distributions

Averaging quantiles

Rough Idea



Black: Averaging probabilities
Grey: Averaging quantiles

Example

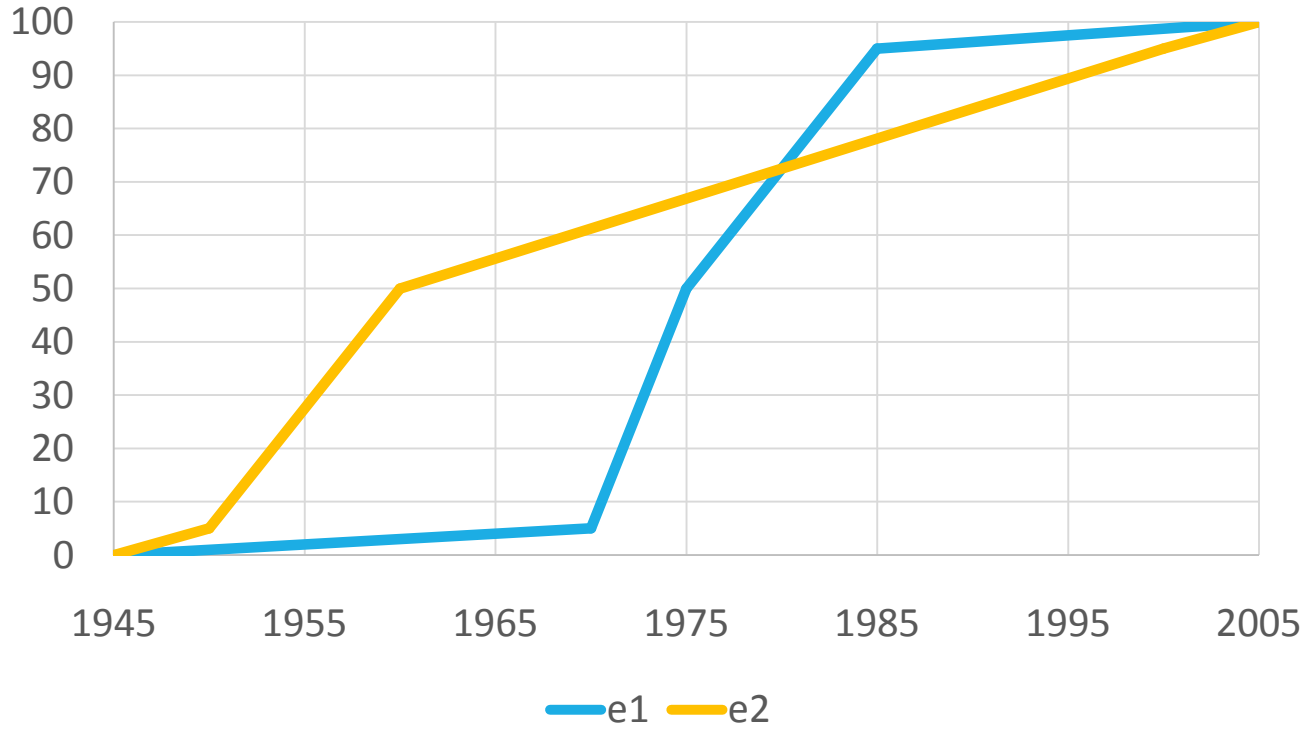
What was the 1946 RAND forecast for

year of first launched satellite?

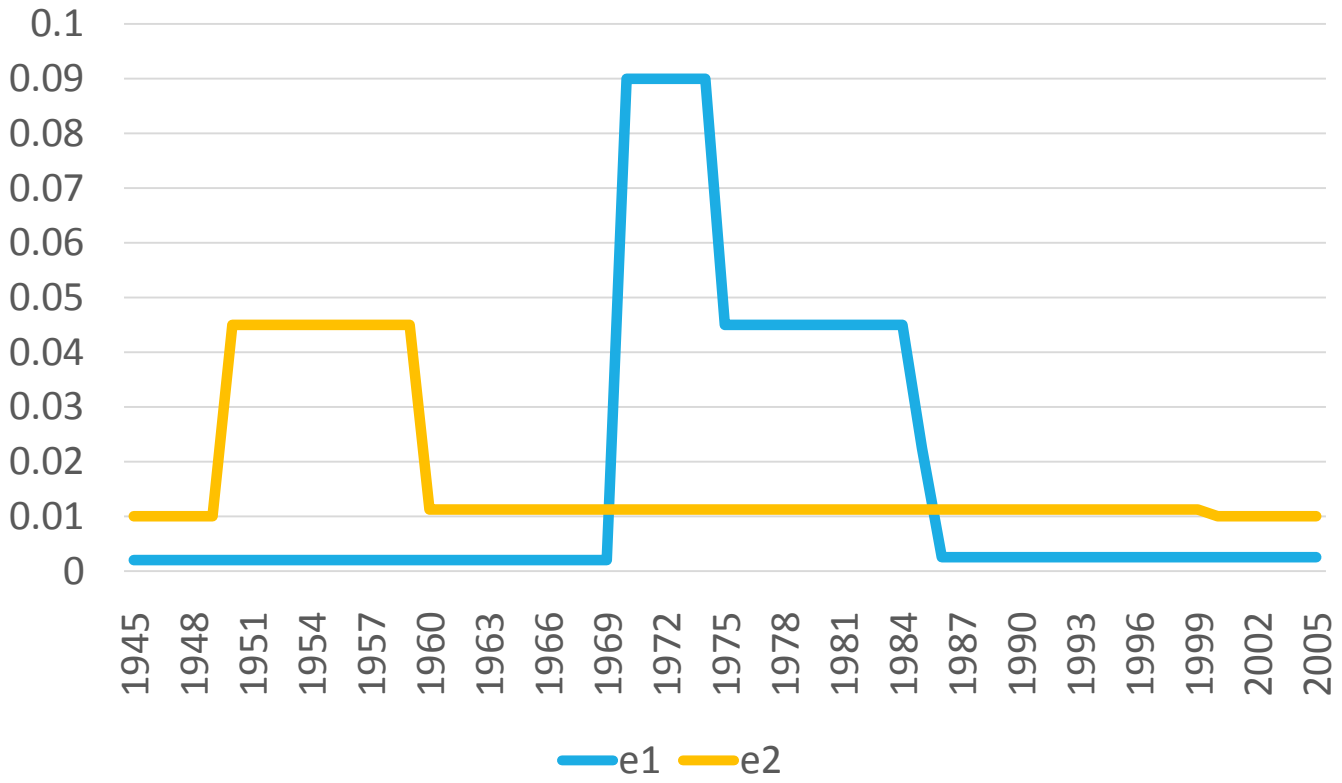
e_1: 5%__1970__, 50%__1975__, 95%__1985__

e_2: 5%__1950__, 50%__1960__, 95%__2000__

Example: cumulative distribution function (cdf) from experts



Example: probability density functions (pdf) from experts



Example: average pdf