

Intervalle de confiance: à quoi ça sert ?



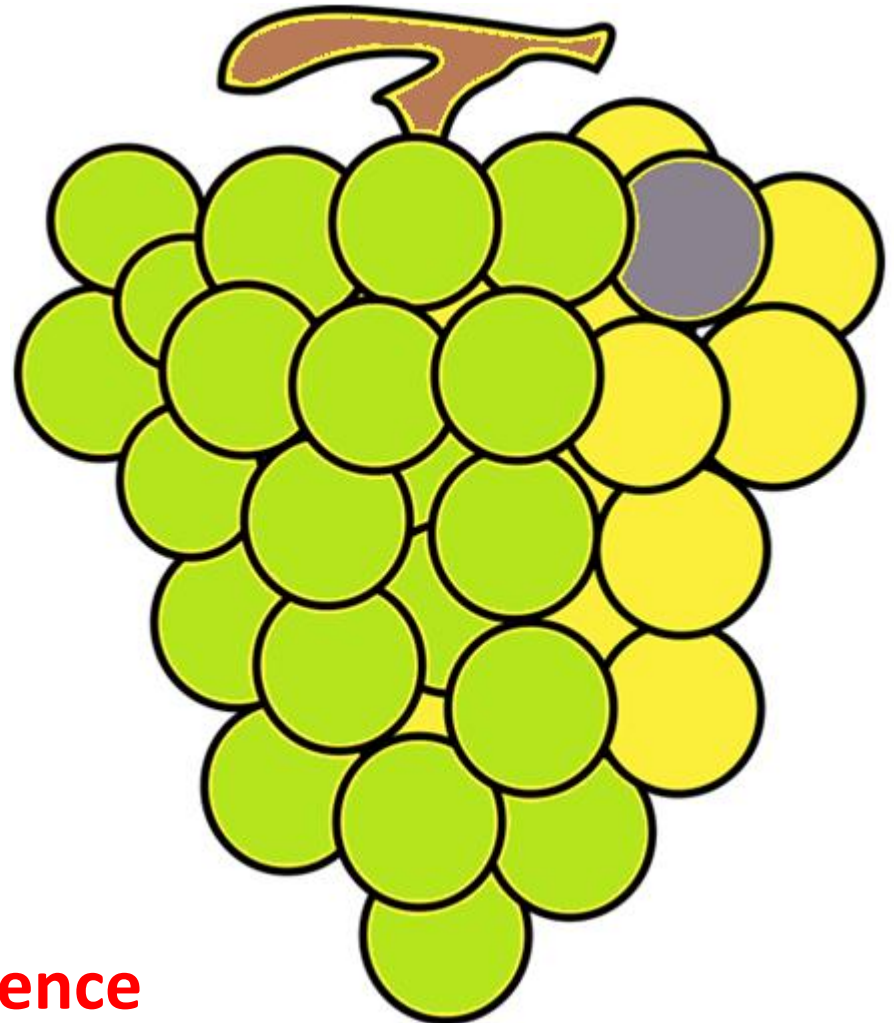
Brindel P., Combescure C.
Service d'épidémiologie clinique
Centre de Recherche Clinique

Plan de la présentation

- Intervalle de confiance à 95%: méthodologie
 - inférence
 - principe de l'intervalle de confiance
 - interprétation de l'intervalle de confiance
 - recommandations
- Revue d'essais cliniques randomisés en oncologie
 - rapportant un effet statistiquement significatif
 - description des intervalles de confiance à 95%

Inférence

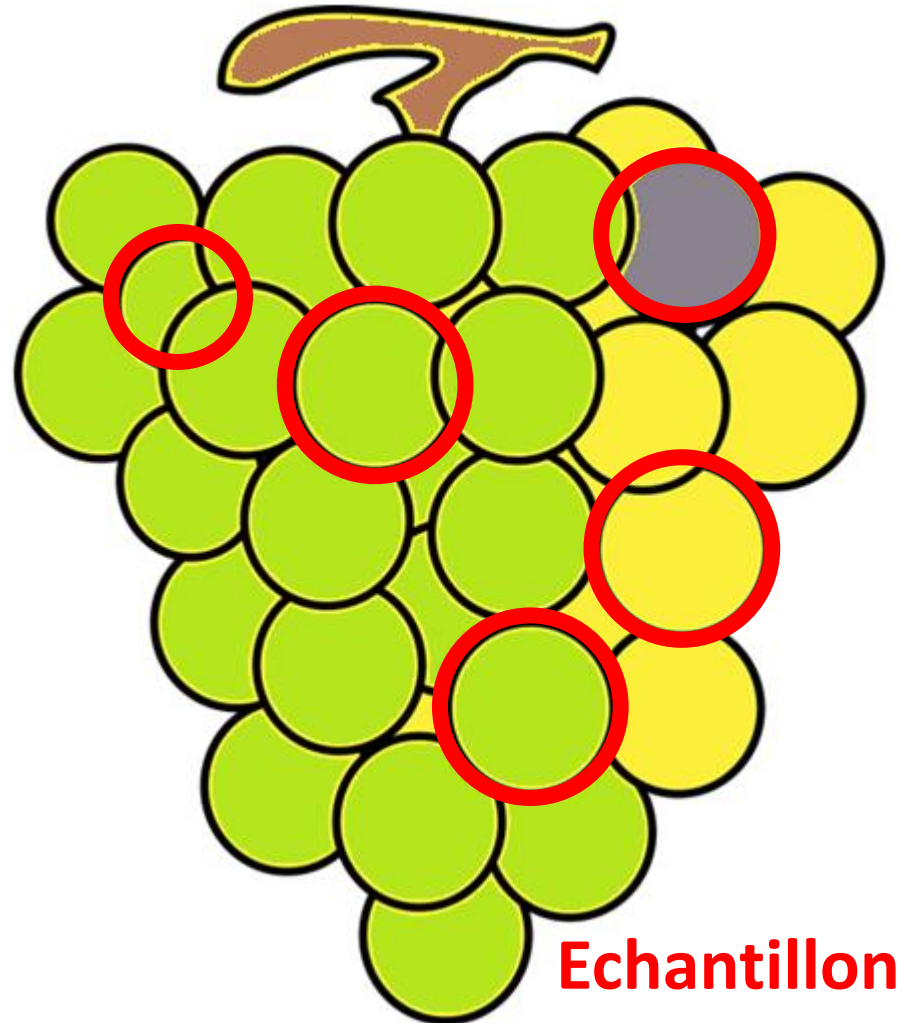
- Comment avoir une idée de la qualité de la grappe de raisin ?
- ◆ Goûter un des raisins :
si le raisin est bon, vous concluez que la grappe est bonne
si le raisin est mauvais, vous concluez que la grappe est mauvaise
- ◆ Vous **généralisez** à l'ensemble de la grappe la qualité du raisin testé



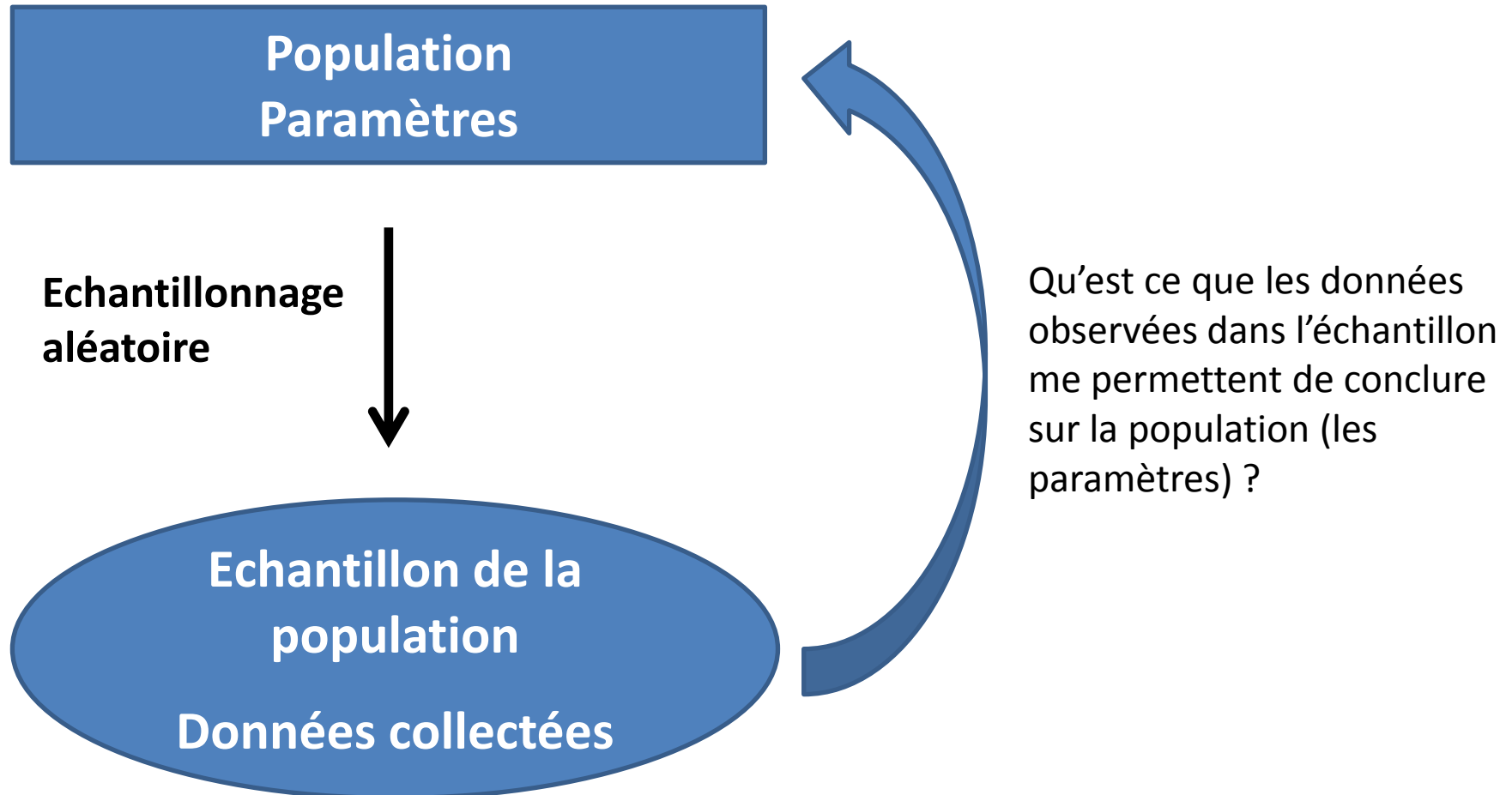
Inférence

Inférence

- Comment sélectionner le raisin à goûter ?
- ◆ Sélection subjective ?
risque de goûter surtout les jolis raisins => inférence incorrecte
- ◆ Sélection **aléatoire**
protège d'une sélection subjective
- ◆ Plusieurs raisins: **échantillon**



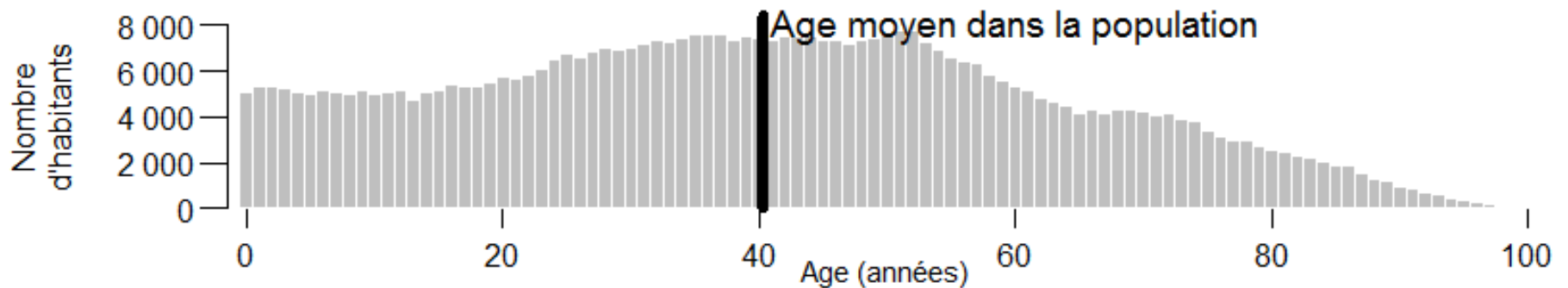
Inférence statistique



Exemple: estimation de l'âge moyen

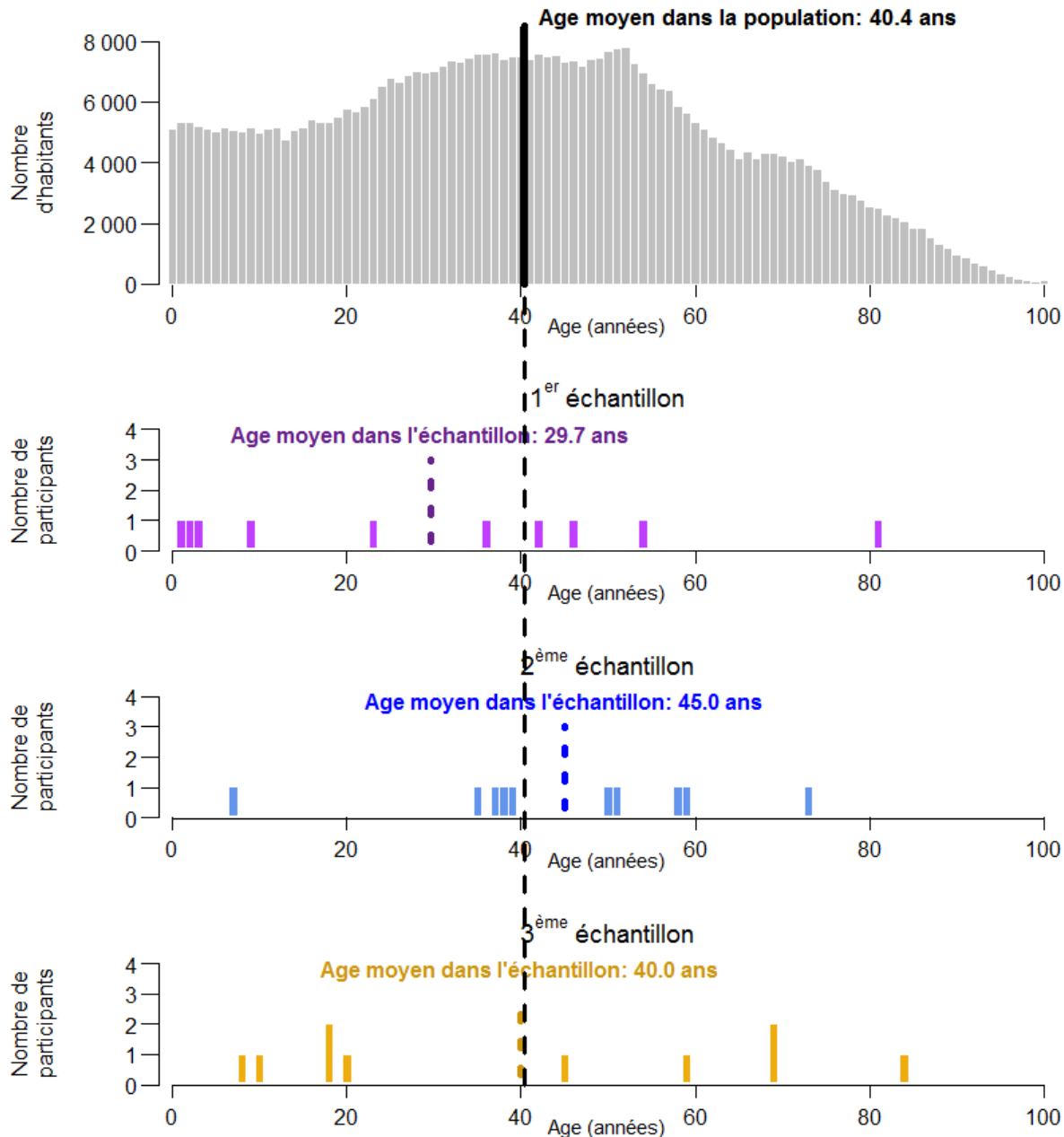
- Source des données: OFS, année 2016
- Données de **toute** la population du canton de Genève
- **Paramètre**: âge moyen 40,4 ans (dans la population)

Distribution de l'âge des habitants du Canton de Genève (2016)



- **Echantillon aléatoire** de cette population
- Age moyen dans l'échantillon => **estimation** de l'âge moyen dans la population
- Comment se comporte l'estimation de l'âge moyen ?

Distribution de l'âge des habitants du Canton de Genève (2016)

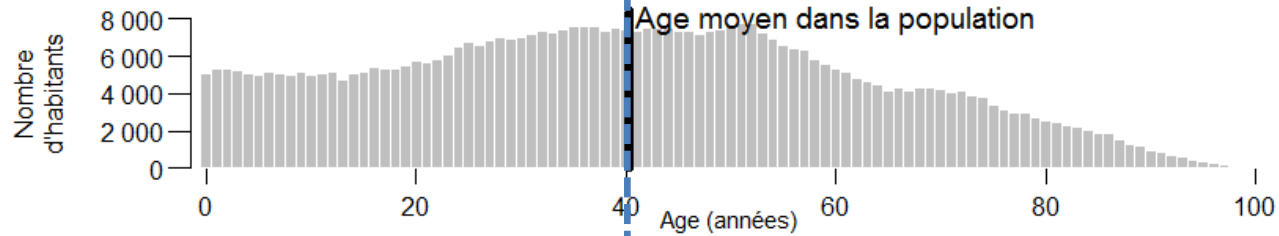


3 échantillons
aléatoires de la
population (n=10)

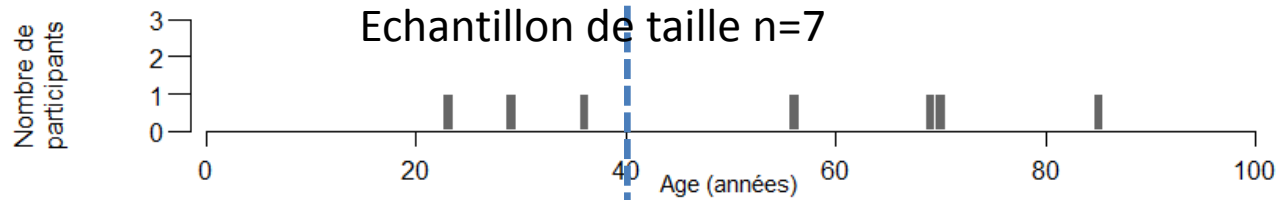
L'estimation de l'âge
moyen dans la
population varie d'un
échantillon à l'autre

Cette variabilité est
causée par l'aléa de
l'échantillonnage

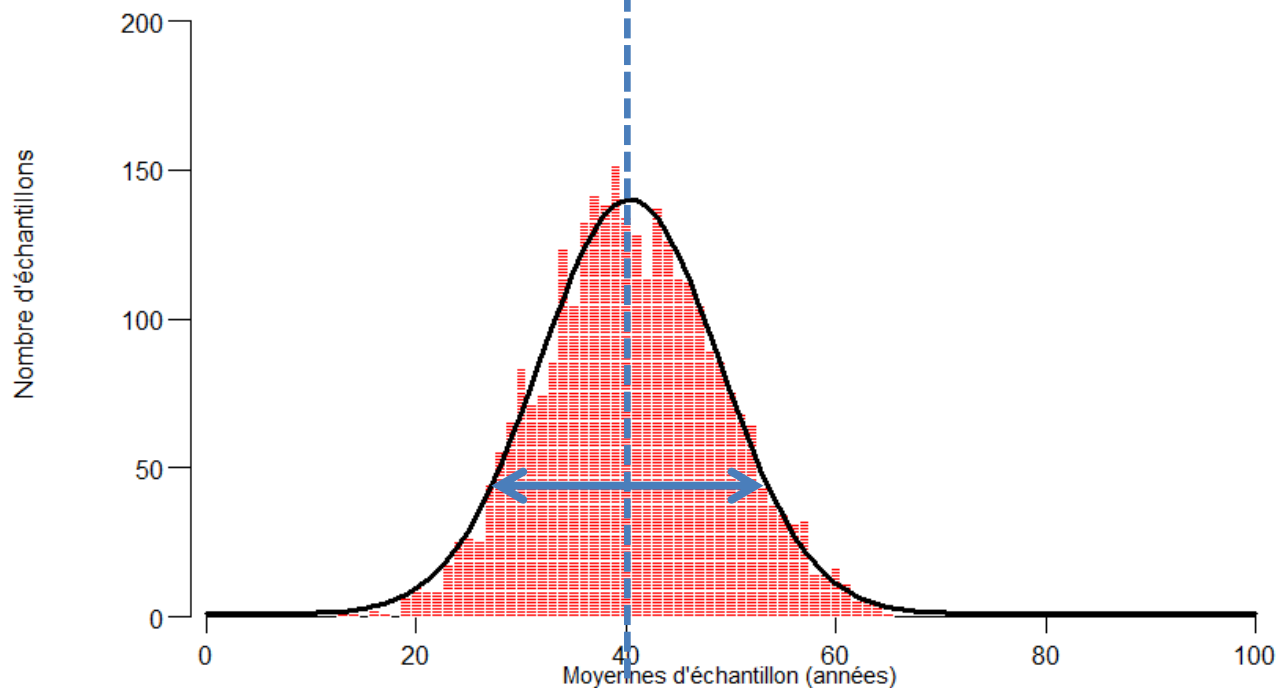
Distribution de l'âge des habitants du Canton de Genève (2016)



Les moyennes d'échantillon sont centrées sur la valeur du paramètre dans la population



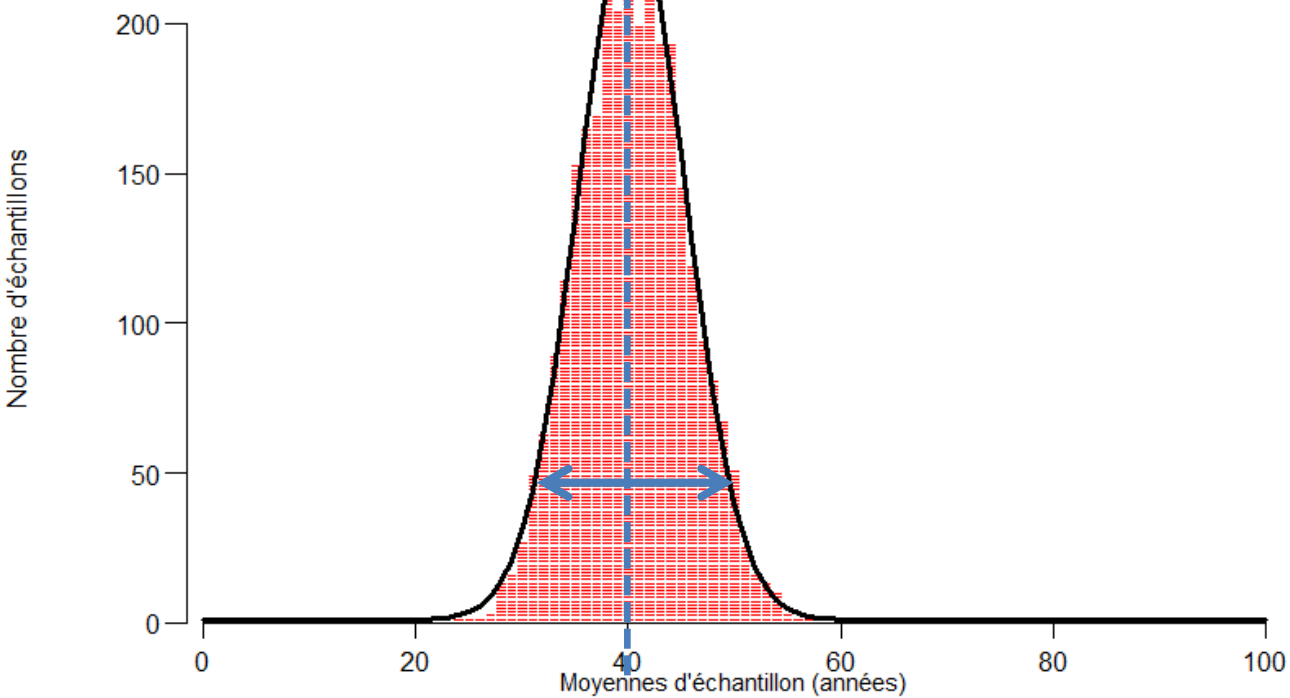
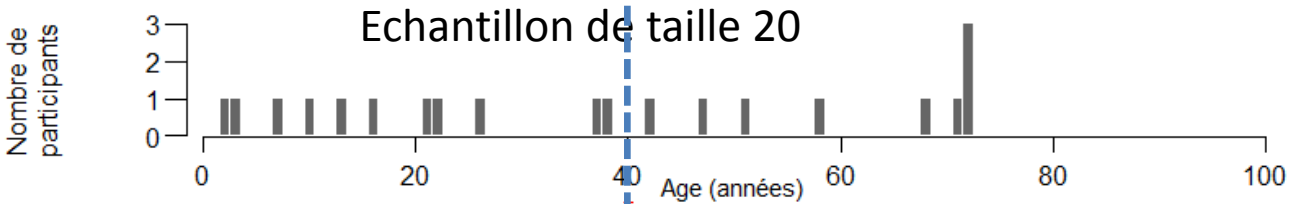
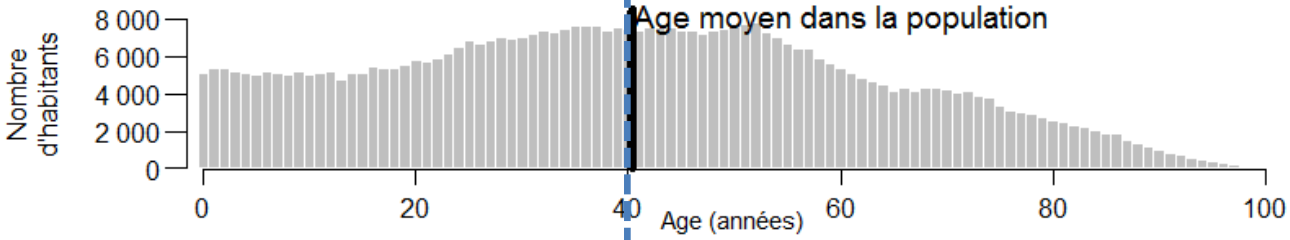
Les moyennes d'échantillon sont variables



Les moyennes d'échantillon ont une distribution en cloche

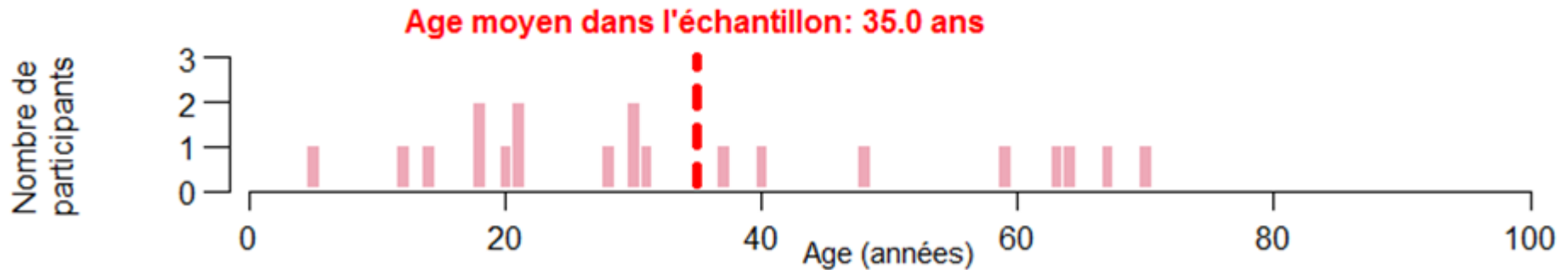
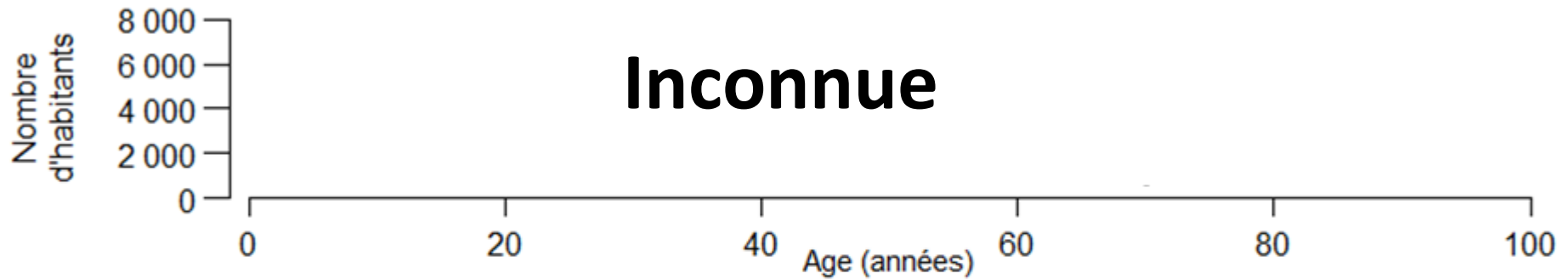
Vrai même si la variable n'est pas gaussienne dans la population (à condition que n soit suffisamment grand)

Distribution de l'âge des habitants du Canton de Genève (2016)



La variabilité des moyennes d'échantillon diminue lorsque la taille d'échantillon augmente

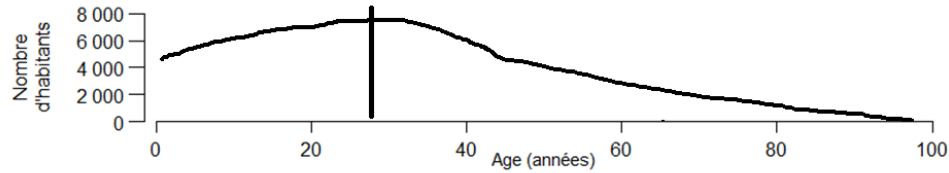
Distribution de l'âge des habitants du Canton de Genève (2016)



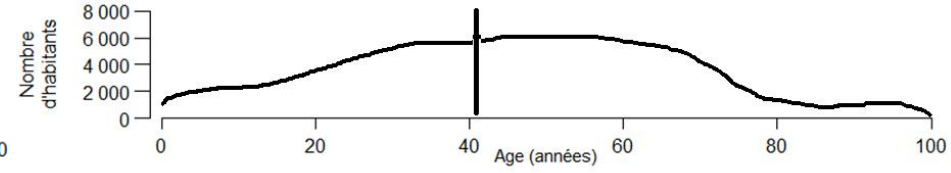
En général, il n'est pas possible de connaître la valeur du paramètre dans la population ou les données de toute la population

Comment savoir dans quelle situation je suis ?

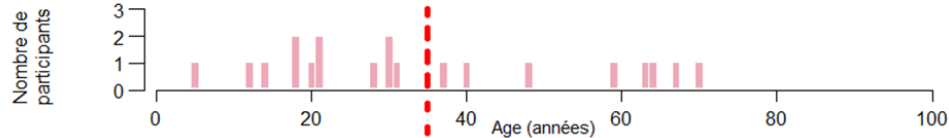
Distribution de l'âge des habitants du Canton de Genève (2016)



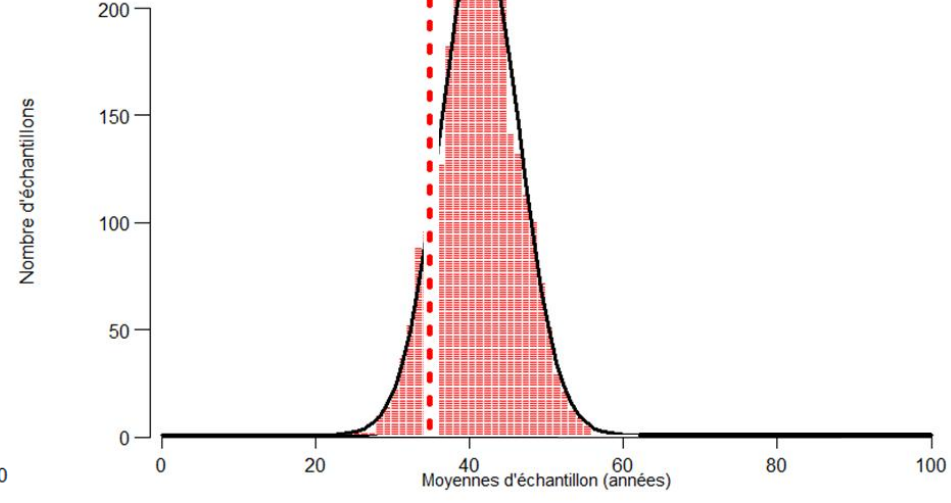
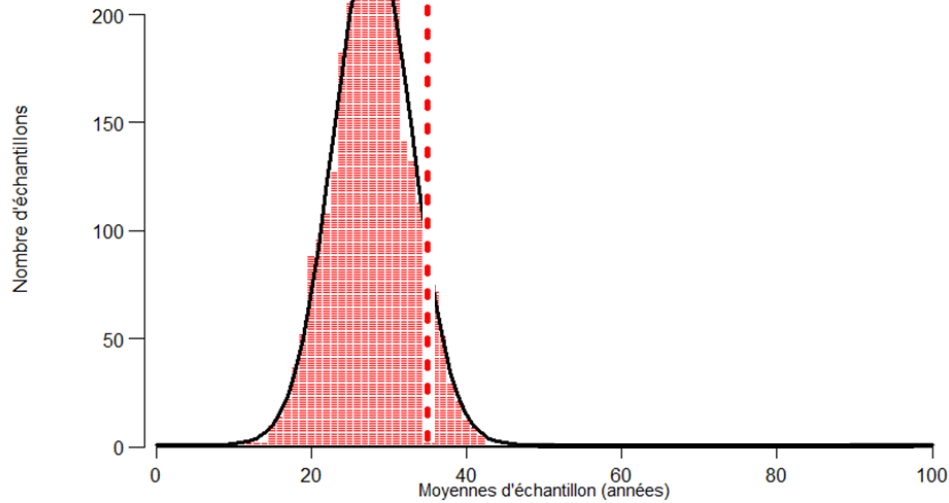
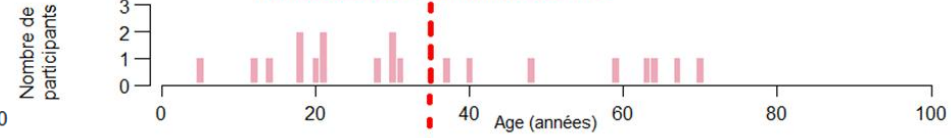
Distribution de l'âge des habitants du Canton de Genève (2016)



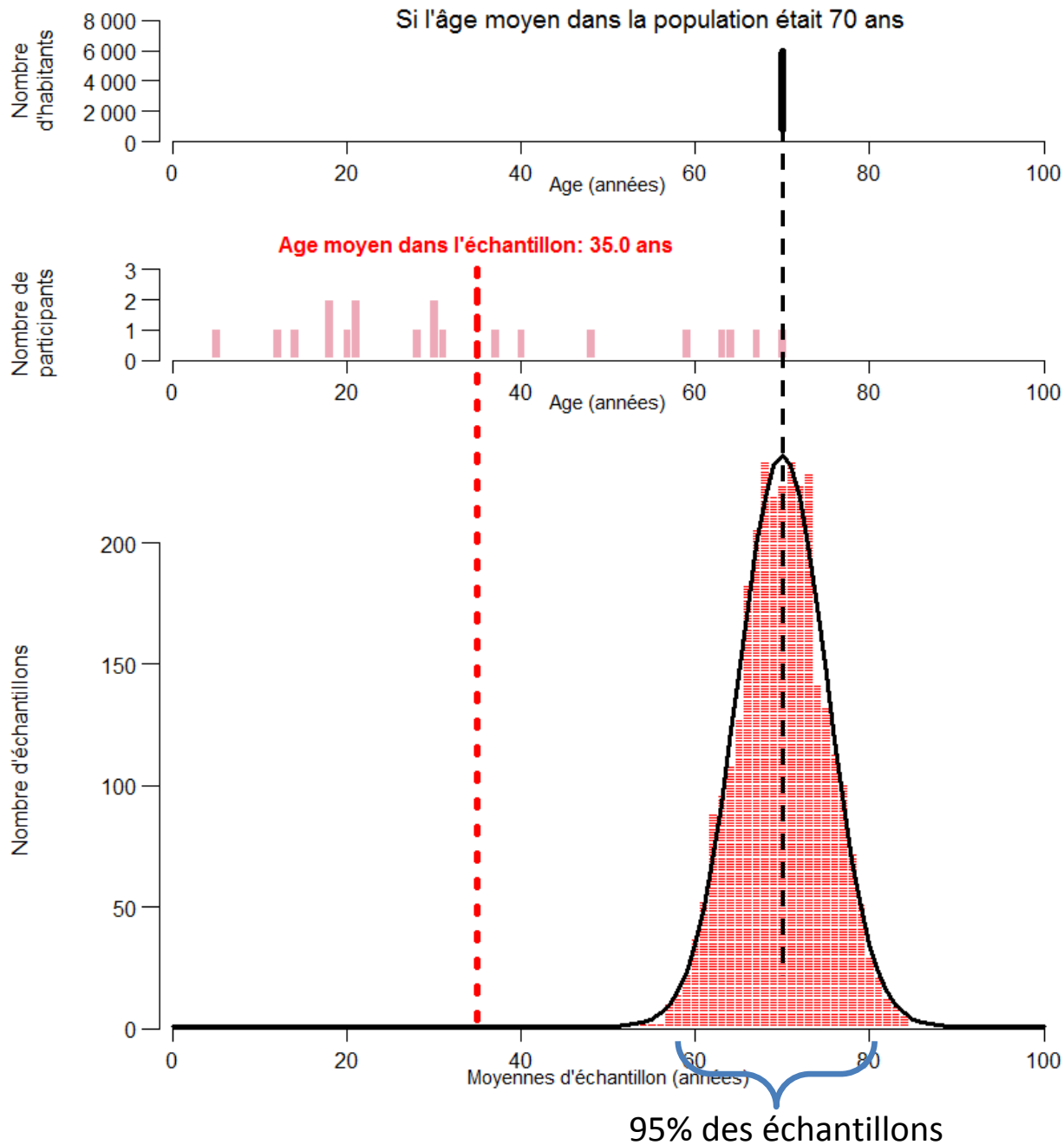
Age moyen dans l'échantillon: 35.0 ans



Age moyen dans l'échantillon: 35.0 ans

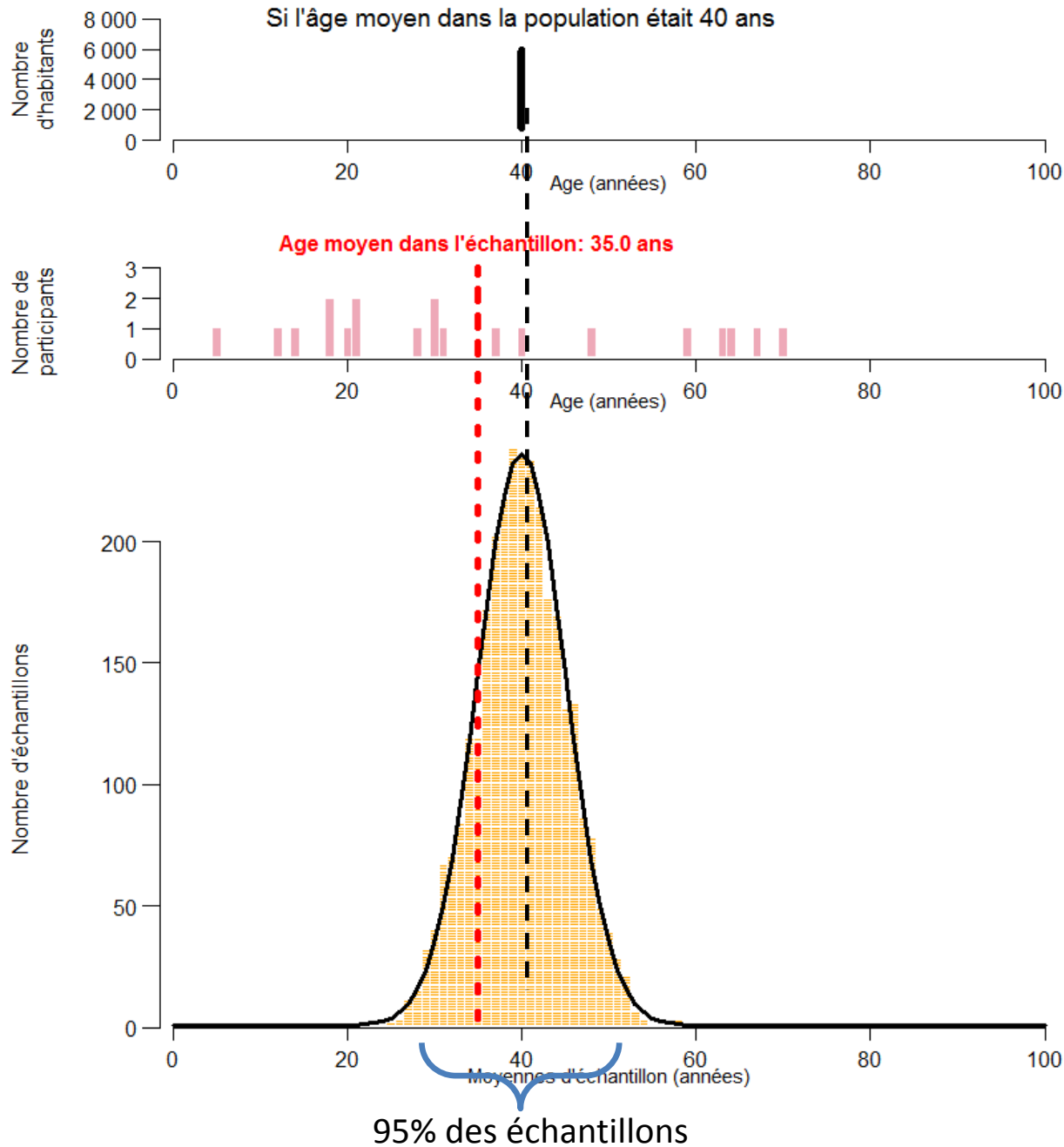


Distribution de l'âge des habitants du Canton de Genève (2016)



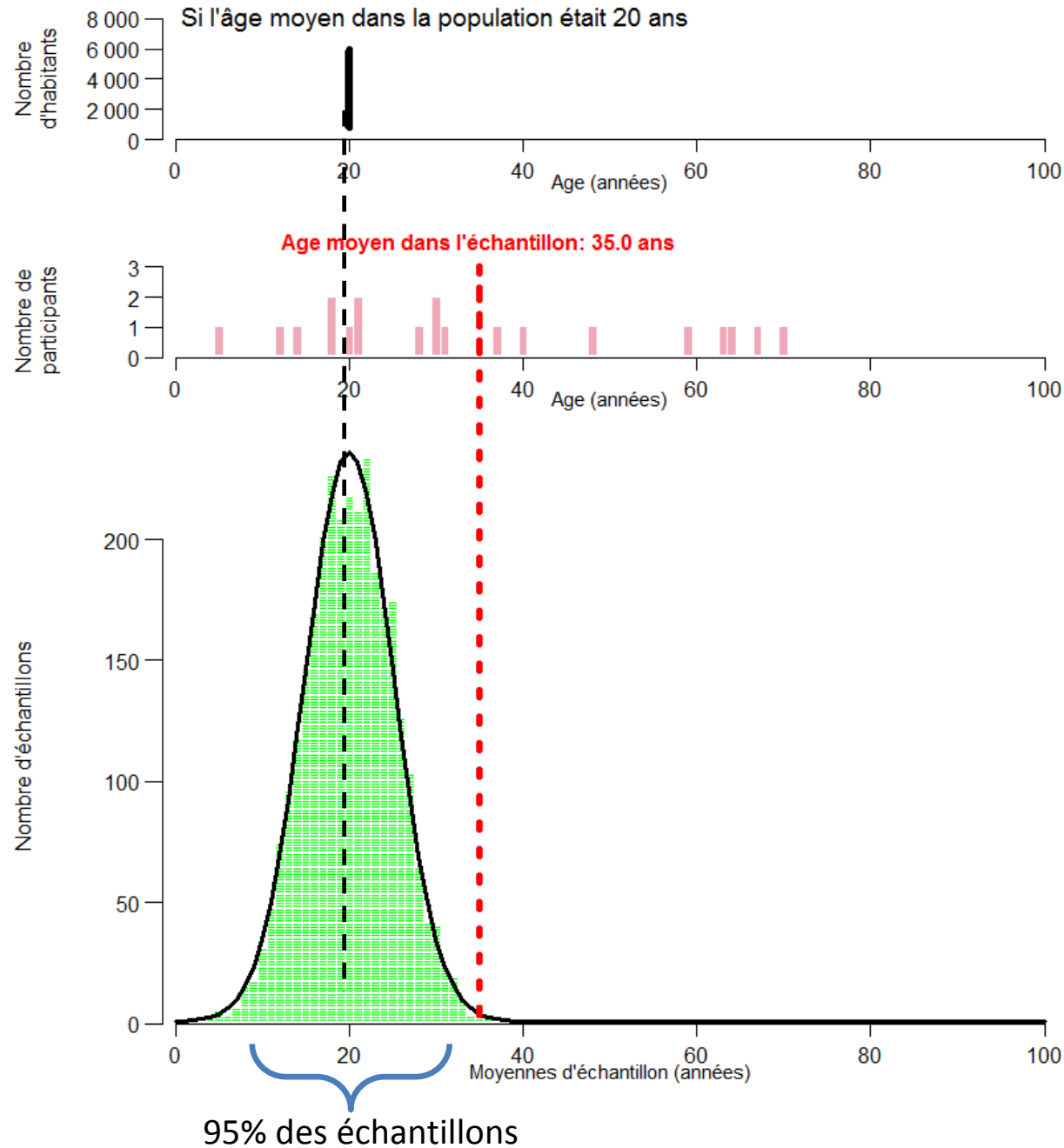
Un âge moyen de **70 ans** **dans la population** n'est pas compatible avec les données observés dans l'échantillon car: l'âge moyen **dans l'échantillon** est en dehors de l'intervalle contenant 95% des échantillons si l'âge moyen était **70 ans** dans la population

Distribution de l'âge des habitants du Canton de Genève (2016)

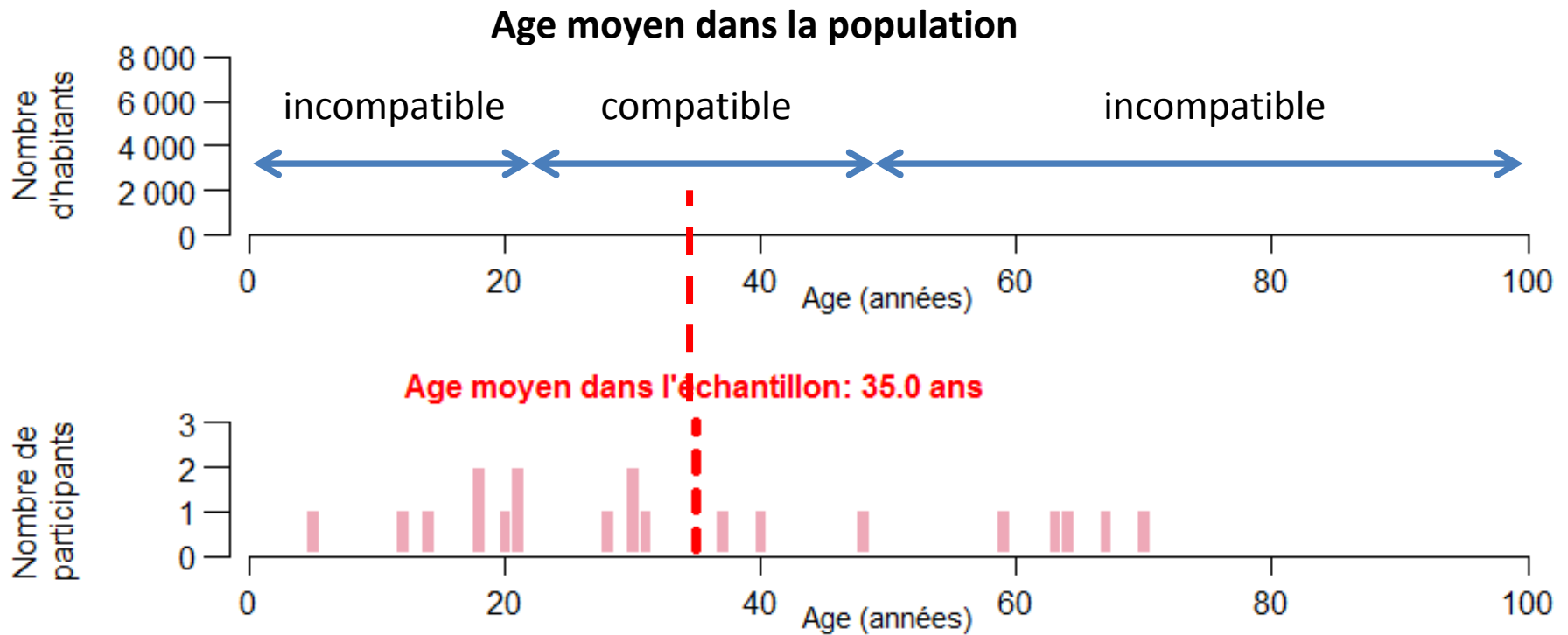


Un âge moyen de **40 ans** **dans la population** est compatible avec les données observés dans l'échantillon car l'âge moyen **dans l'échantillon** est dans l'intervalle contenant 95% des échantillons si l'âge moyen était **40 ans** dans la population

Distribution de l'âge des habitants du Canton de Genève (2016)



Un âge moyen de **20 ans** **dans la population** n'est pas compatible avec les données observés dans l'échantillon car l'âge moyen **dans l'échantillon** est en dehors de l'intervalle contenant 95% des échantillons si l'âge moyen était **20 ans** dans la population



Intervalle de confiance:

ensemble des valeurs du paramètre (dans la population) qui sont compatibles avec les données observées dans l'échantillon

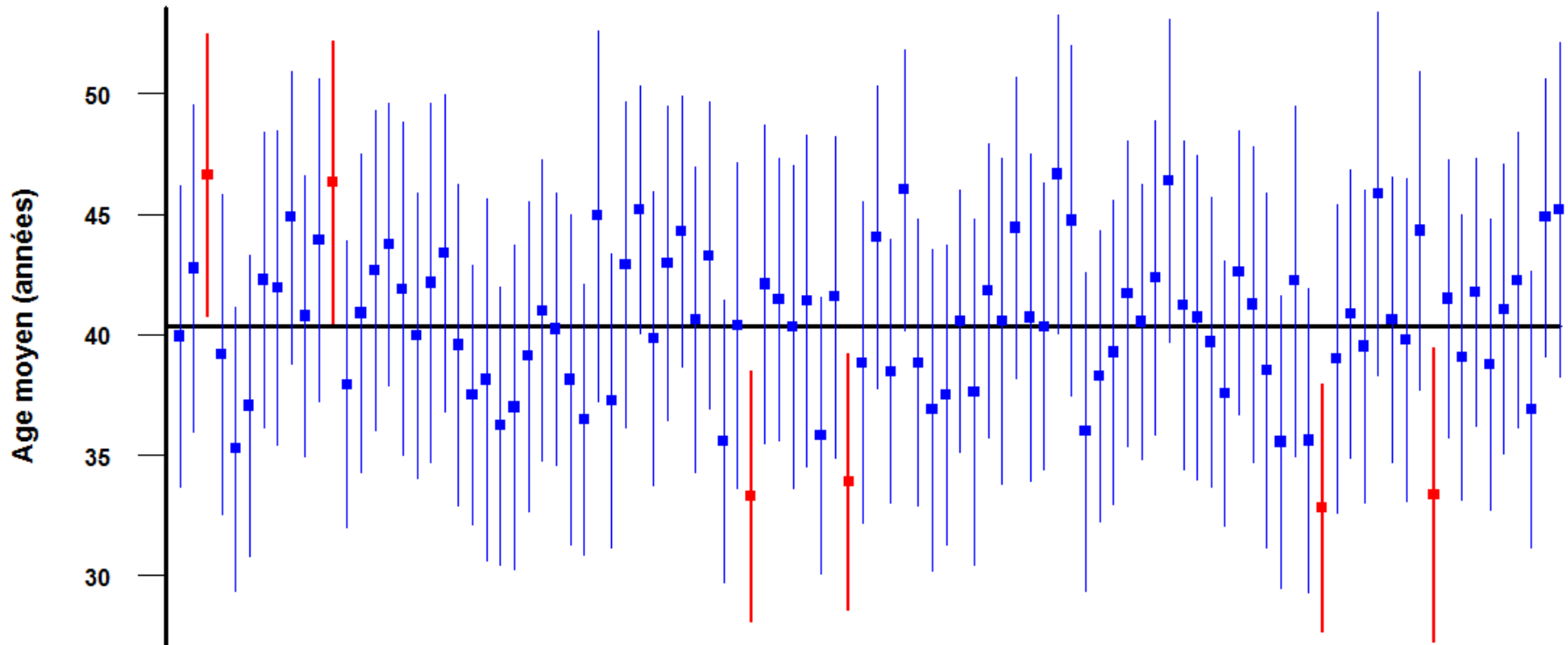
Intervalles de confiance à 95%

- Propriétés:
 - toujours centré sur l'estimation
 - largeur diminue lorsque le nombre d'observations augmente
- L'intervalle de confiance à 95% existe pour toute estimation (moyenne, proportion, taux d'incidence, odds ratio, risque relatif, différence de moyennes, hazard ratio...)
- Formules pour une moyenne et une proportion :

$$\text{moyenne estimée} \pm 2 \frac{\text{écart type}}{\sqrt{\text{taille d'échantillon}}}$$

$$\text{proportion estimée} \pm 2 \sqrt{\frac{\text{proportion estimée} (1 - \text{proportion estimée})}{\text{taille d'échantillon}}}$$

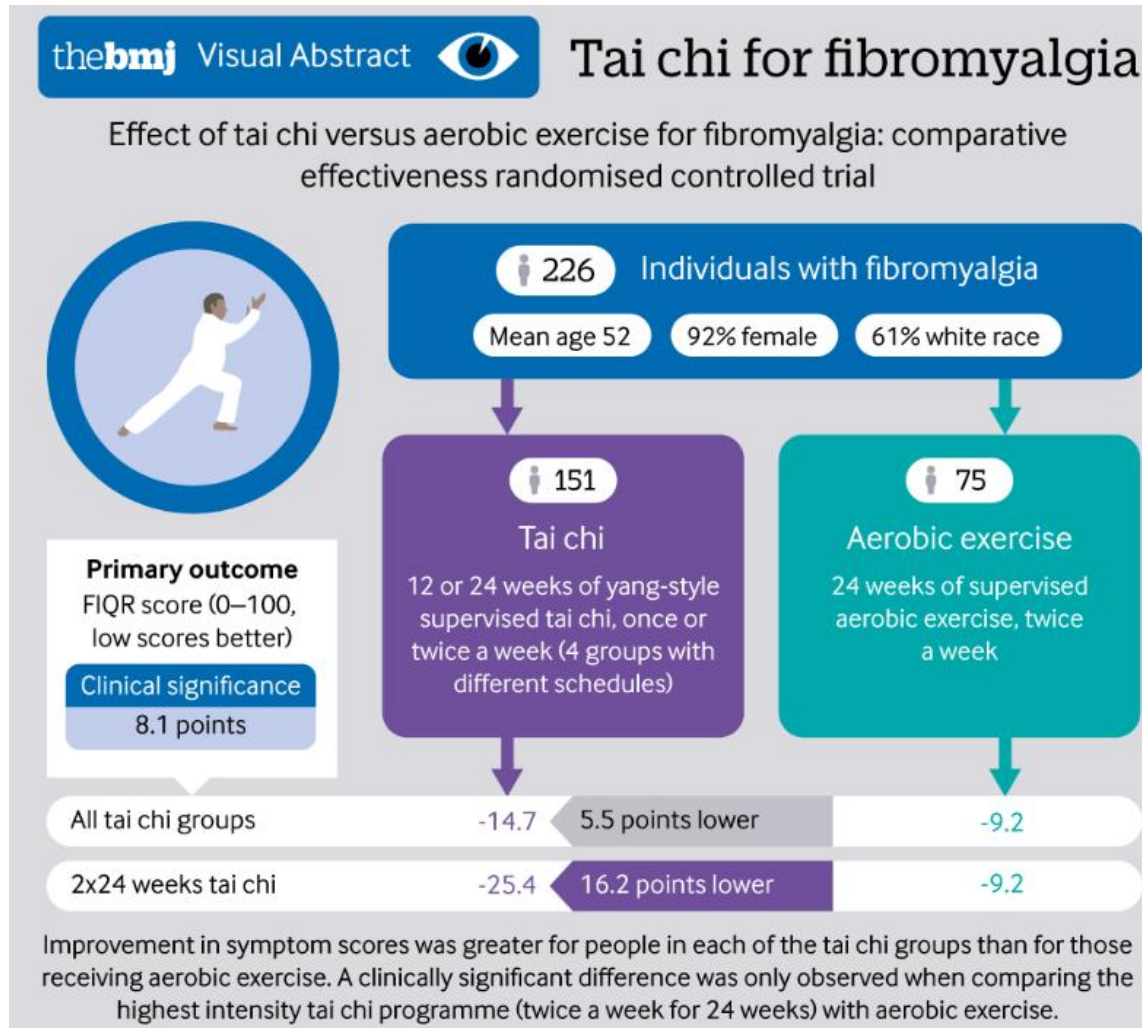
100 échantillons de taille 50



- Sur un grand nombre d'échantillons aléatoires de la même population, il est attendu que dans 5% des échantillons, l'IC95% ne contient pas la vraie valeur du paramètre (quelle que soit la taille d'échantillon)
- En pratique, le chercheur a un seul échantillon et il ne sait pas si l'intervalle de confiance contient ou non la vraie valeur

Effect of tai chi versus aerobic exercise for fibromyalgia: comparative effectiveness randomised controlled trial

Chenchen Wang,¹ Christopher H Schmid,² Roger A Fielding,³ William F Harvey,¹ Kieran F Reid,³ Lori Lyn Price,⁴ Jeffrey B Driban,¹ Robert Kalish,⁵ Ramel Rones,⁶ Timothy McAlindon¹



DESIGN

Prospective, randomized, 52 week, single blind comparative effectiveness trial.

SETTING

Urban tertiary care academic hospital in the United States between March 2012 and September 2016.

PARTICIPANTS

226 adults with fibromyalgia (as defined by the American College of Rheumatology 1990 and 2010 criteria) were included in the intention to treat analyses: 151 were assigned to one of four tai chi groups and 75 to an aerobic exercise group.

Primary outcome

The primary outcome was the change in revised FIQR total score (range 0-100, with higher scores indicating greater impact of symptoms) from baseline to the 24 week visit.²⁸

Therefore, we estimated the minimal clinically important difference of the FIQR score based on the established minimal clinically important difference of the FIQ score (14% or 8.1 units).^{40 19}

Table 3 | Between group differences at all weeks*

Variables	Aerobic exercise v tai chi groups combined		Aerobic exercise 2×24 weeks v tai chi 2×24 weeks	
	Mean (95% CI)	P value	Mean (95% CI)	P value
FIQR score (range 0-100)†:				
Week 12	5.4 (0.6 to 10.1)	0.03	10.9 (3.4 to 18.5)	0.005
Week 24	5.5 (0.6 to 10.4)	0.03	16.2 (8.7 to 23.6)	<0.001
Week 52	2.7 (-2.3 to 7.7)	0.29	11.1 (2.7 to 19.6)	0.01

L'effet estimé est +5.5 en faveur du tai chi (sur le score FIQR)

Significativité statistique: l'hypothèse d'absence d'effet dans la population (« vraie différence de moyenne = 0) est rejetée

L'intervalle de confiance à 95% va de 0.6 à 10.4: les données de l'échantillon permettent de dire que, en réalité, l'effet du tai chi peut être 0.6 ou 10.4.

L'intervalle de confiance à 95% contient 8.1 (la plus petite différence cliniquement pertinente): les données de l'échantillon ne permettent pas d'exclure que le véritable effet est plus faible que la plus petite différence cliniquement pertinente.

Table 3 | Between group differences at all weeks*

Variables	Aerobic exercise v tai chi groups combined		Aerobic exercise 2×24 weeks v tai chi 2×24 weeks	
	Mean (95% CI)	P value	Mean (95% CI)	P value
FIQR score (range 0-100)†:				
Week 12	5.4 (0.6 to 10.1)	0.03	10.9 (3.4 to 18.5)	0.005
Week 24	5.5 (0.6 to 10.4)	0.03	16.2 (8.7 to 23.6)	<0.001
Week 52	2.7 (-2.3 to 7.7)	0.29	11.1 (2.7 to 19.6)	0.01

L'effet estimé est +16.2 en faveur du tai chi (**2x24 semaines**)

Significativité statistique: l'hypothèse d'absence d'effet dans la population (« vraie différence de moyenne = 0) est rejetée

L'intervalle de confiance à 95% va de 8.7 à 23.6: les données de l'échantillon permettent de dire que, en réalité, l'effet du tai chi peut être 8.7 ou 23.6.

L'intervalle de confiance à 95% ne contient pas la valeur 8.1 (la plus petite différence cliniquement pertinente): les données de l'échantillon permettent d'exclure que le véritable effet est plus faible que la plus petite différence cliniquement pertinente.

Les recommandations

The **CONSORT** statement (CONsolidated Standards Of Reporting Trials)

Tableau 1 (Suite)

Section/Topic	Item N°	Checklist item
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its <u>precision (such as 95 % confidence interval)</u>

STROBE Statement—checklist of items that should be included in reports of observational studies

Results

Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their <u>precision (eg, 95% confidence interval)</u> . Make clear which confounders were adjusted for and why they were included
--------------	----	--

PRISMA 2009 Checklist

RESULTS

Synthesis of results	21	Present results of each meta-analysis done, <u>including confidence intervals</u> and measures of consistency.
----------------------	----	--

7. How large was the treatment effect?

HINT: Consider

- what outcomes were measured
- Is the primary outcome clearly specified
- what results were found for each outcome

8. How precise was the estimate of the treatment effect?

HINT: Consider

- what are the confidence limits

Consider:

- *if a confidence interval were reported. Would your decision about whether or not to use this intervention be the same at the upper confidence limit as at the lower confidence limit?*
- *if a p-value is reported where confidence intervals are unavailable*

Conclusion

- L'intervalle de confiance à 95% est l'ensemble des valeurs du paramètre compatibles avec les données de l'échantillon
- L'IC 95%:
 - mesure la précision de l'estimation
 - capte l'incertitude liée à l'aléa de l'échantillonnage
 - ne mesure pas l'impact d'éventuel biais sur l'estimation
 - est utile pour tirer des conclusions sur la population
 - n'est pas pertinent pour décrire les données
 - ne contient pas toujours la valeur du paramètre estimé (mais vous ne saurez pas si c'est le cas ou non dans votre étude!)
- Un chercheur doit
 - le comprendre
 - le rapporter
 - l'interpréter



Limites de l'intervalle de confiance

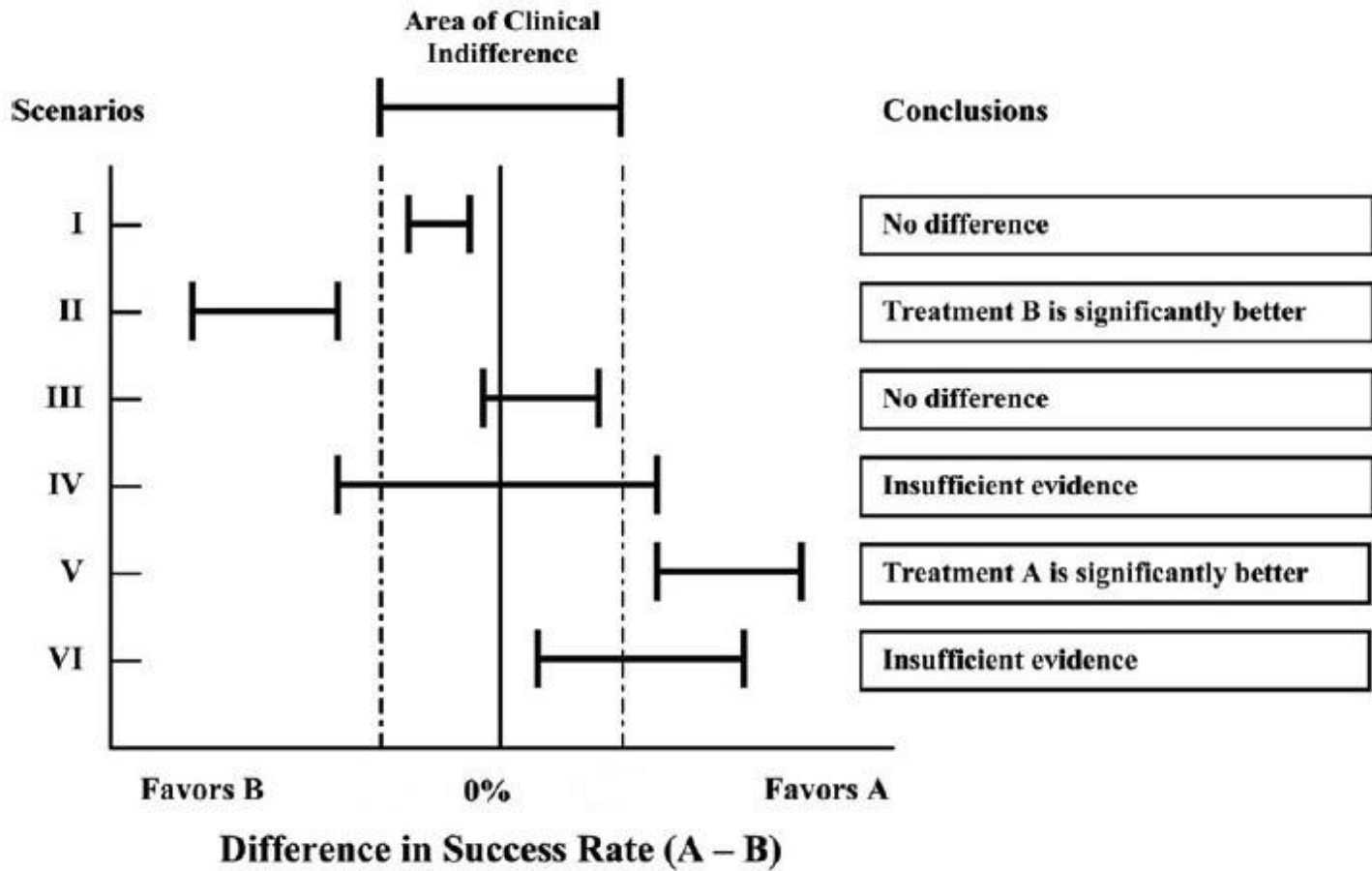
Exemple à partir d'une revue de la littérature
des résultats statistiquement significatifs d'essai
contrôlés randomisés en oncologie

Pauline Brindel

Colloque du 18 juin 2018

Bénéfices statistiquement significatifs
en terme de survie globale dans les
RCT en oncologie:

Sont-ils potentiellement non
pertinents cliniquement ?

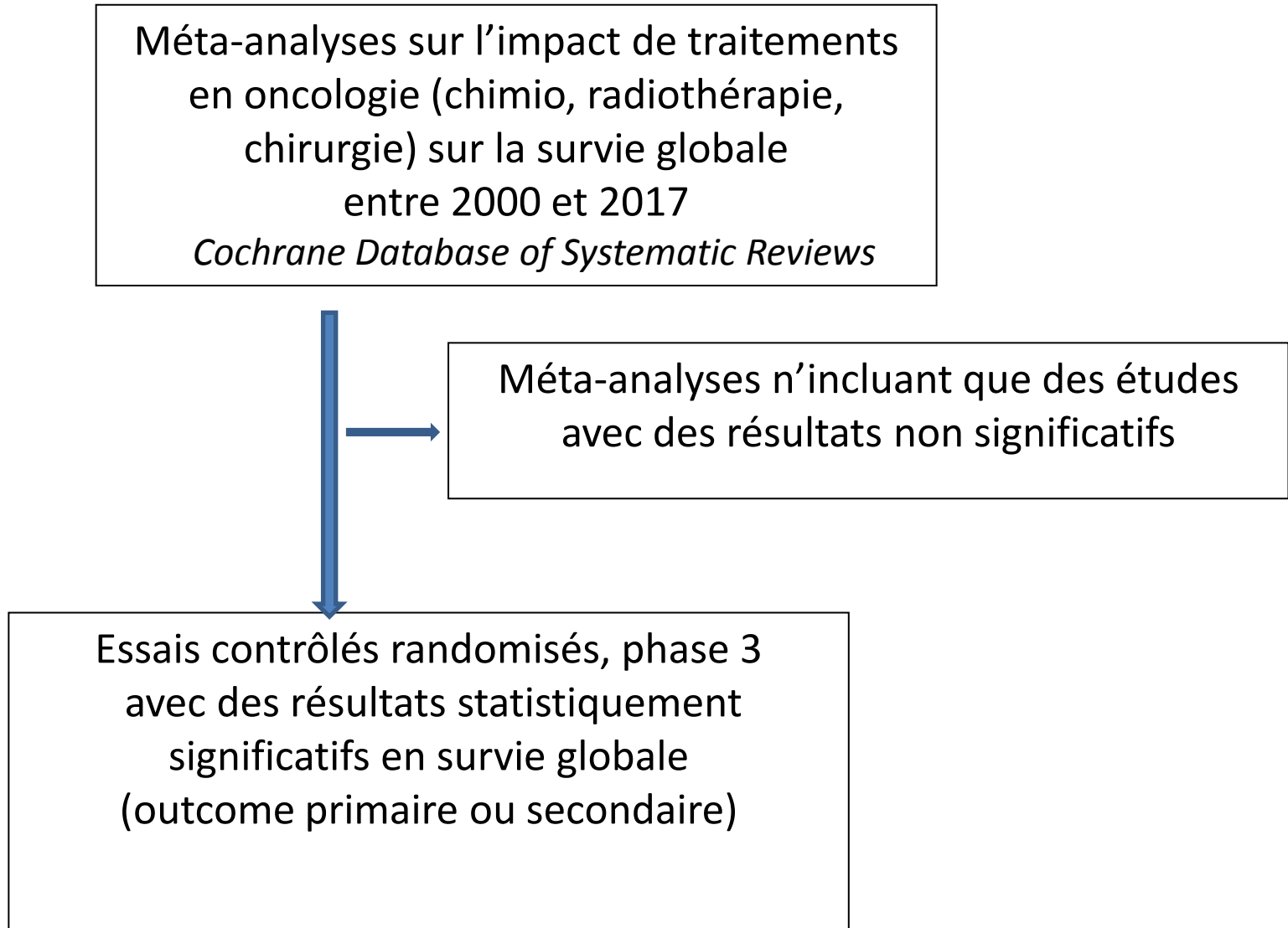


D'après : Minimal clinically important difference (MCID)--adding meaning to statistical inference. Chan LS. Am J Public Health. 2013 Nov;103(11):e24-5.

Rationnel

- Parmi des résultats statistiquement significatifs, à quelle fréquence observe-t-on un intervalle de confiance compatible avec une différence qui est potentiellement non *pertinente* cliniquement?

Matériel et méthodes



Matériel et méthodes: données

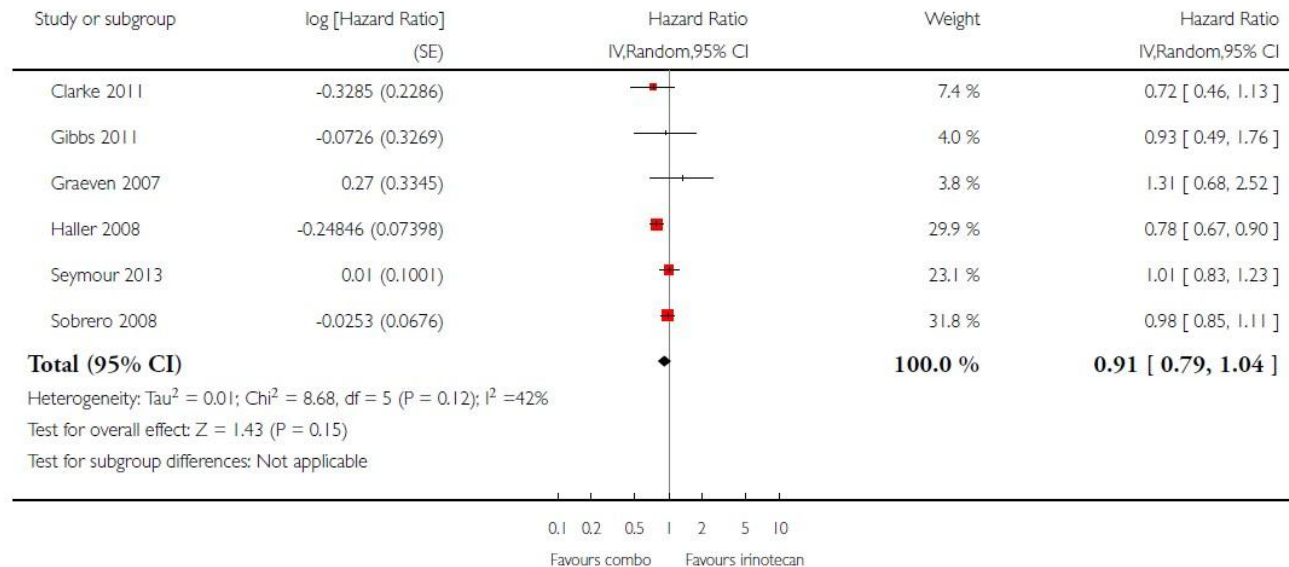
- Pour chaque étude avec un résultat significatif:
 - Point estimate pour la survie globale (outcome primaire ou secondaire): hazard ratio (HR)
 - Intervalle de confiance 95% du hazard ratio pour la survie globale

Analysis 1.5. Comparison 1 All, Outcome 5 OS:irino combination (combo).

Review: Second-line systemic therapy for metastatic colorectal cancer

Comparison: 1 All

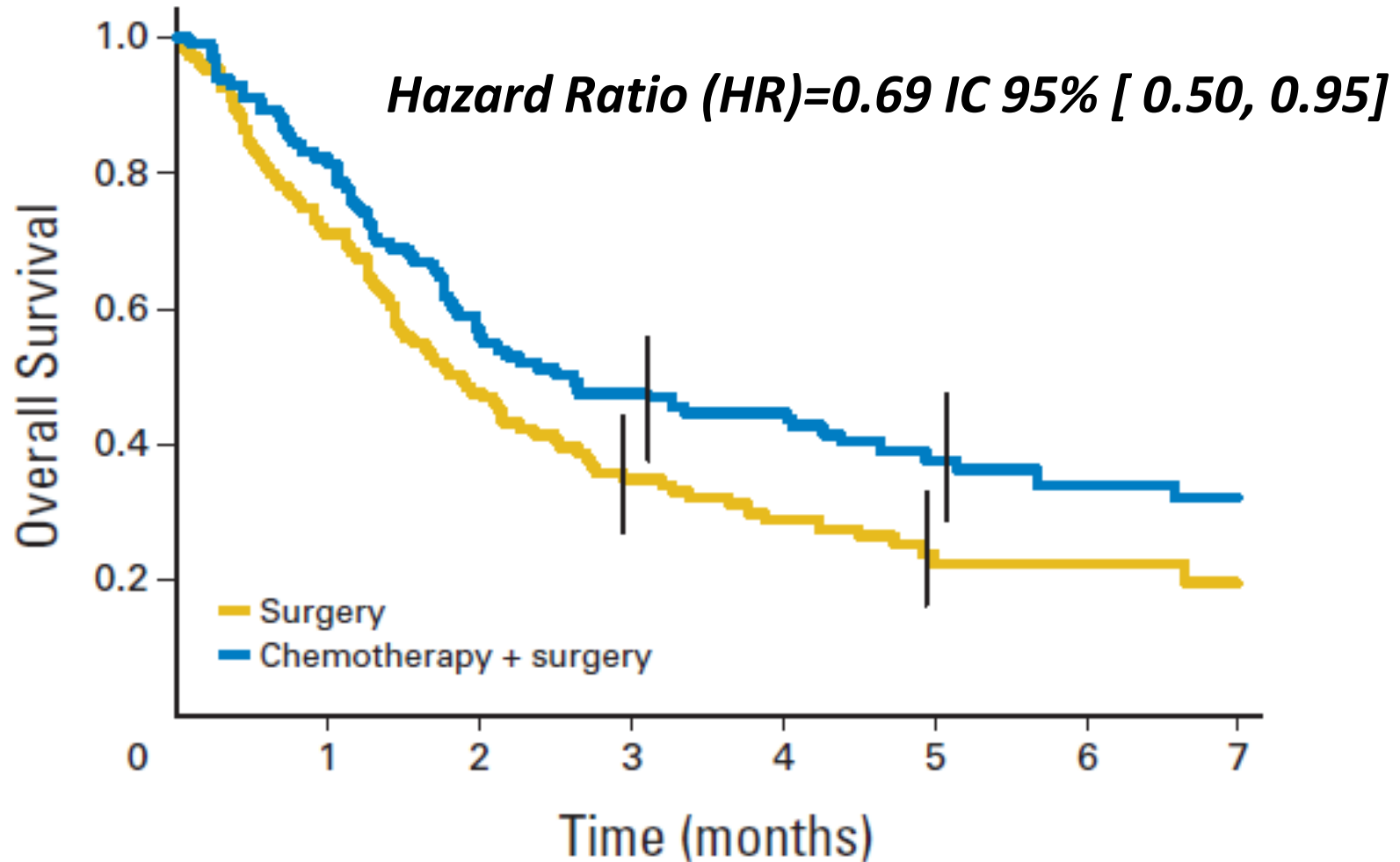
Outcome: 5 OS:irino combination (combo)



Rappel : Analyse de survie

Courbes de Kaplan-Meier de survie:

chimiothérapie + chirurgie (n=113) comparée au groupe contrôle chirurgie (n=111)



Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFCD multicenter phase III trial. Ychou M et al. J Clin Oncol. 2011 May 1;29(13):1715-21

Matériel et méthodes : analyses

Description de la distribution des HR et des limites supérieures des intervalles de confiance.

- Où se situe la borne supérieure de l'IC?
- A quelle fréquence est-elle plus ou moins proche de 1 (hypothèse nulle d'absence de différence ou d'absence de bénéfice)

Résultats

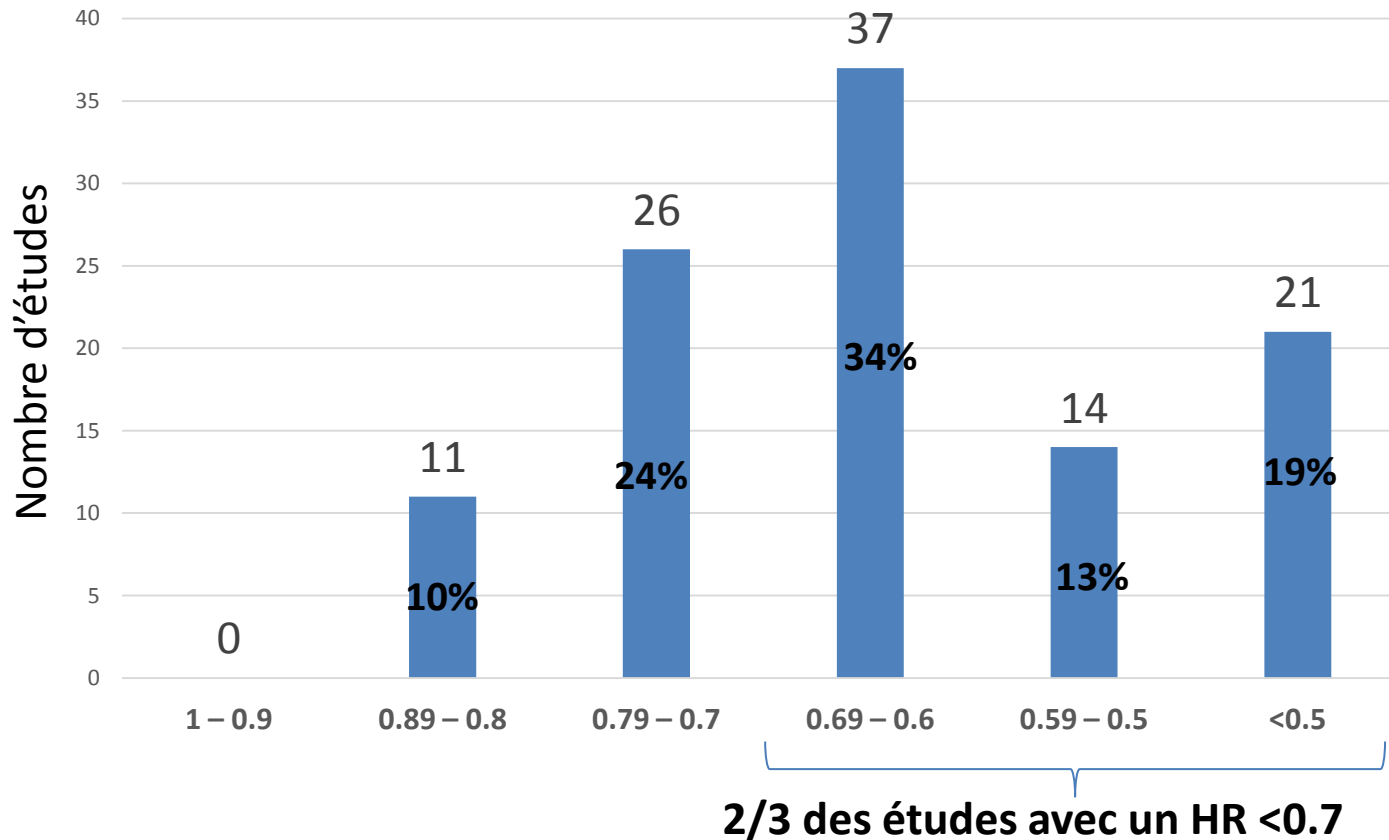
41 revues Cochrane sélectionnées
entre 2000 et 2017
56 méta-analyses
N=451 études



342 Études exclues

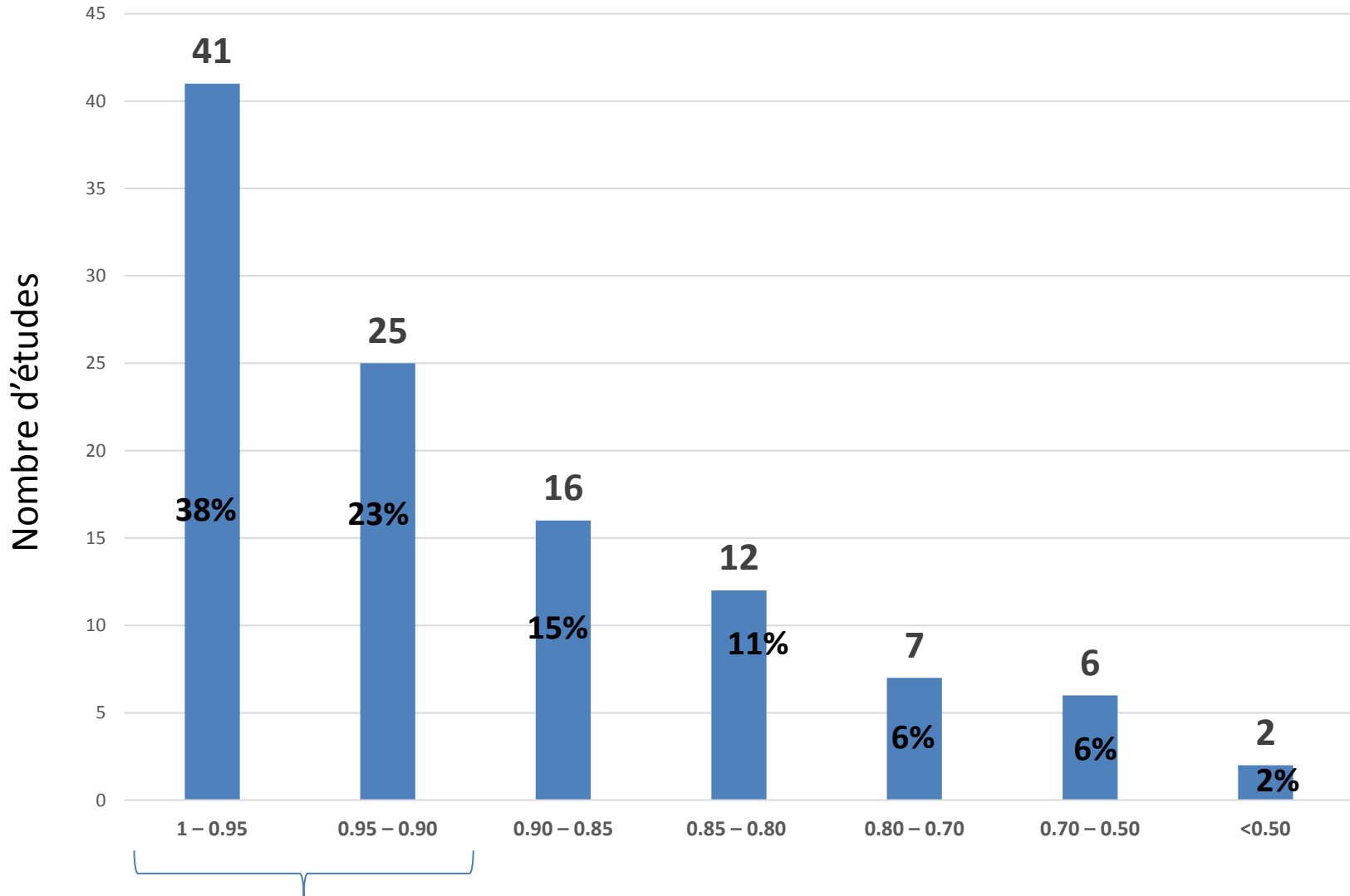
n=109 études avec des résultats montrant
un bénéfice statistiquement significatif de
survie globale

Résultats : Distribution des Hazards Ratios (n=109)



**Gain de survie relative plus important
par rapport au groupe contrôle**

Résultats : Distribution des limites supérieures des intervalles de confiance (n=109)



66 études, soit 61%, avec une limite supérieure de l'IC entre 0.90 et 1

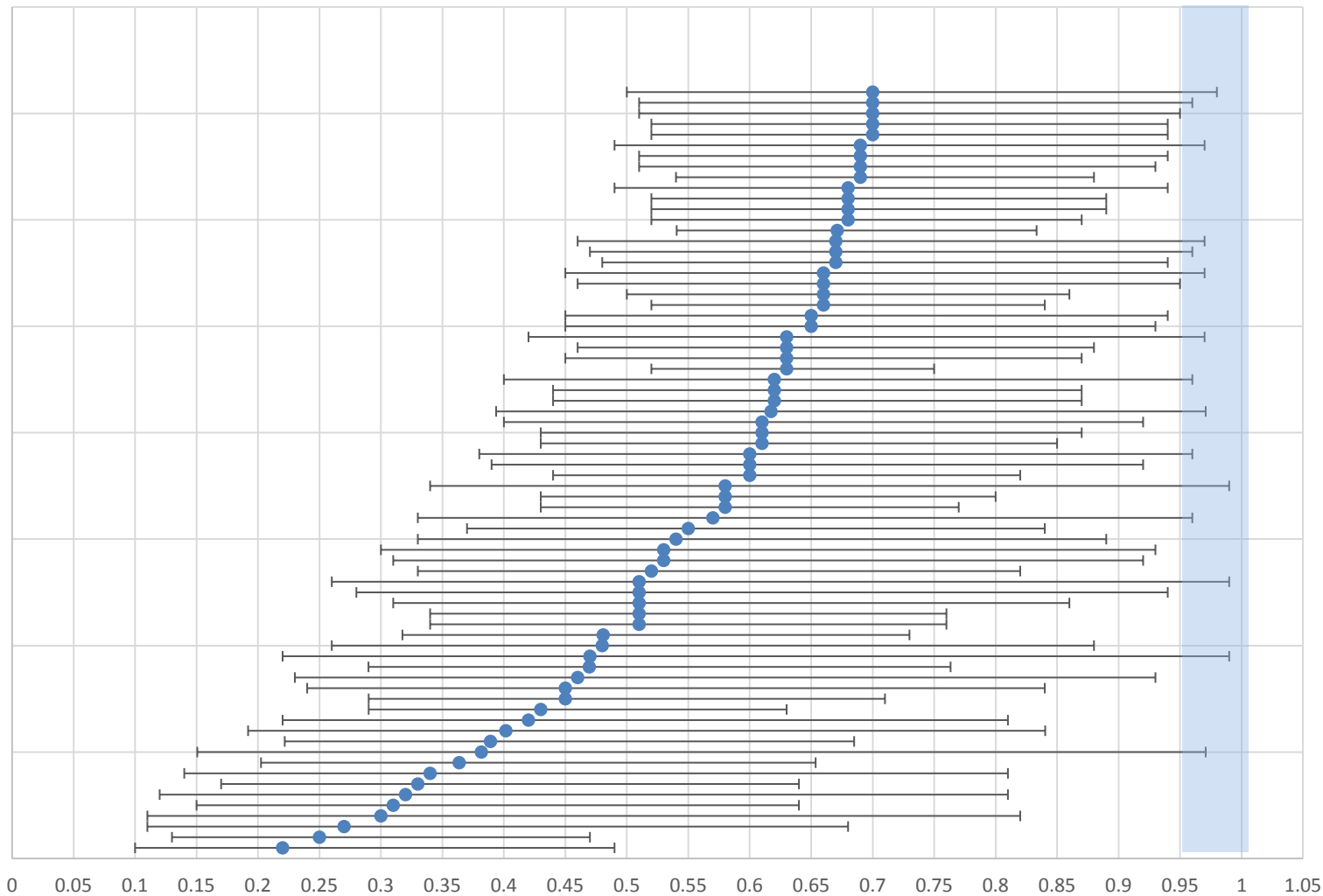
Résultats

Tableau 2: Distribution des “point estimates” et des limites supérieures de l’IC (n=109)

IC 95%	Point estimate (Hazard Ratio)						Total
	1 – 0.9	0.89 – 0.8	0.79 – 0.7	0.69 – 0.6	0.59 – 0.5	<0.5	
Limite supérieure							
1 – 0.95		10	14	12	3	2	41
0.95 – 0.90		1	10	10	3	1	25
0.90 – 0.85			2	11	2	1	16
0.85 – 0.80				3	3	6	12
0.80 – 0.70				1	3	3	7
0.70 – 0.50						6	6
<0.50						2	2
Total	0	11	26	37	14	21	109

2/3 des études avec un point estimate <0.7

Hazard ratios (HR) et intervalles de confiance, pour HR<0.7 (n=72)



24% des IC indiquent des résultats plausibles avec une réduction relative de mortalité de 5% dans le groupe intervention par rapport au groupe contrôle

Discussion

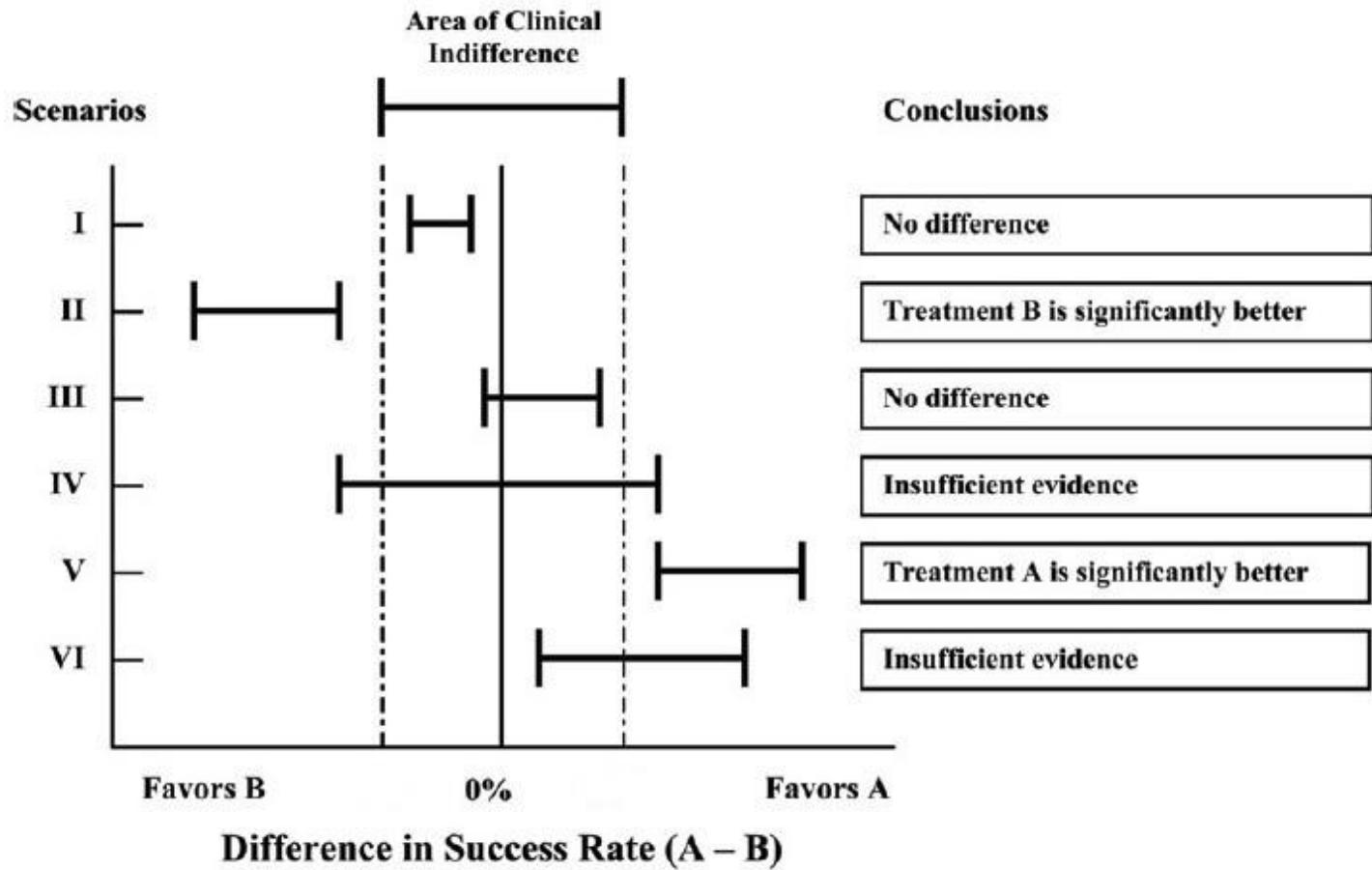
- Parmi bénéfiques de survies décrits comme statistiquement significatifs dans notre échantillon:
 - *réduction relative de mortalité de 30% et plus pour 2/3 des études.*
- Mais,
 - *pour 61% des essais, la limite supérieure de l'IC suggère une réduction relative de mortalité de 10% par rapport au groupe témoin*
 - *décalage entre ce qui est rapporté par HR et IC*
 - *Dépend de l'interprétation du résultat et où on place le curseur pour décider de ce qui est pertinent cliniquement*

Discussion

- Limites de notre étude:
 - pas accès à toutes les explications sur la différence cliniquement attendue dans le calcul de taille d'échantillon des études sélectionnées

Discussion

- Explications pour ces résultats
 - (1) effet réel est plus faible que celui dont on a fait l'hypothèse
 - (2) problème de recrutement, trop optimiste dans calcul de taille d'échantillon
 - (3) On ne peut exclure que dans un certain nombre de cas, l'IC sera proche de l'absence de différence, même si cette différence existe dans la population, lié aux fluctuations d'échantillonnage.



D'après : Minimal clinically important difference (MCID)--adding meaning to statistical inference. Chan LS. Am J Public Health. 2013 Nov;103(11):e24-5.

Conclusion et perspectives

- L'interprétation du résultat en clinique devrait prendre en compte l'incertitude du bénéfice reflétée par l'intervalle de confiance ainsi que la borne de l'intervalle de confiance la plus proche de l'absence de bénéfice.

Conclusion

- Pour aller plus loin...
 - Proposition: designs d'essai avec puissance suffisante pour:
 - détecter un bénéfice donné de survie
 - et exclure de faibles bénéfices de survie dont la pertinence clinique est questionnable

Références

Sedgwick P, Joeques K. Interpreting hazard ratios. BMJ. 2015 Aug 28;351:h4631. doi: 10.1136/bmj.h4631.

Chan LS. Minimal clinically important difference (MCID)--adding meaning to statistical inference. Am J Public Health. 2013 Nov;103(11):e24-5. doi: 10.2105/AJPH.2013.301580. Epub 2013 Sep 12.

<http://www.cochranelibrary.com/about/explanations-for-cochrane-summary-of-findings-sof-tables.html>