MANCHESTER 1824
The University of Manchester

Challenges in RCTs of in vitro fertilisation: an estimands perspective

Jack Wilkinson

Centre for Biostatistics

University of Manchester

@jd_wilko

A confession to PSI

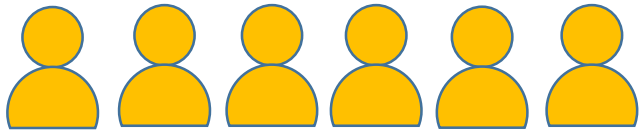
- I do quite a lot of work in the field of assisted reproduction, which includes in vitro fertilisation (IVF).
- I have been quite slow to embrace the estimands framework.
- Topic: Reframing IVF RCT issues using estimand perspective.
- Motivation: force myself to think these things through.

MANCHESTER 1824
The University of Manchester

Outline

- Analysis of neonatal outcomes in IVF trials

We’re going to consider this with reference to the ICH E9 addendum on estimands



Stimulation of ovaries

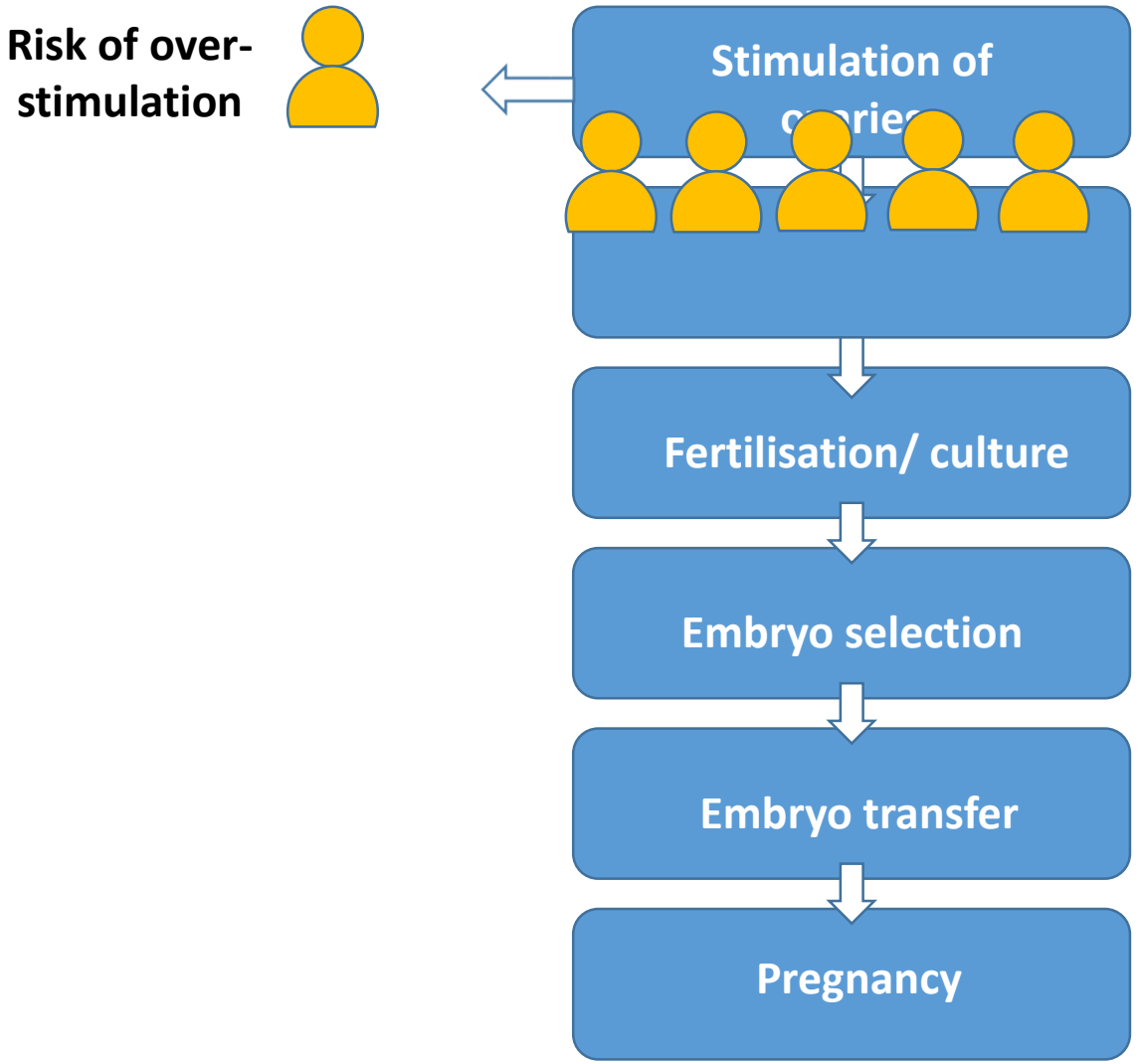
Egg retrieval

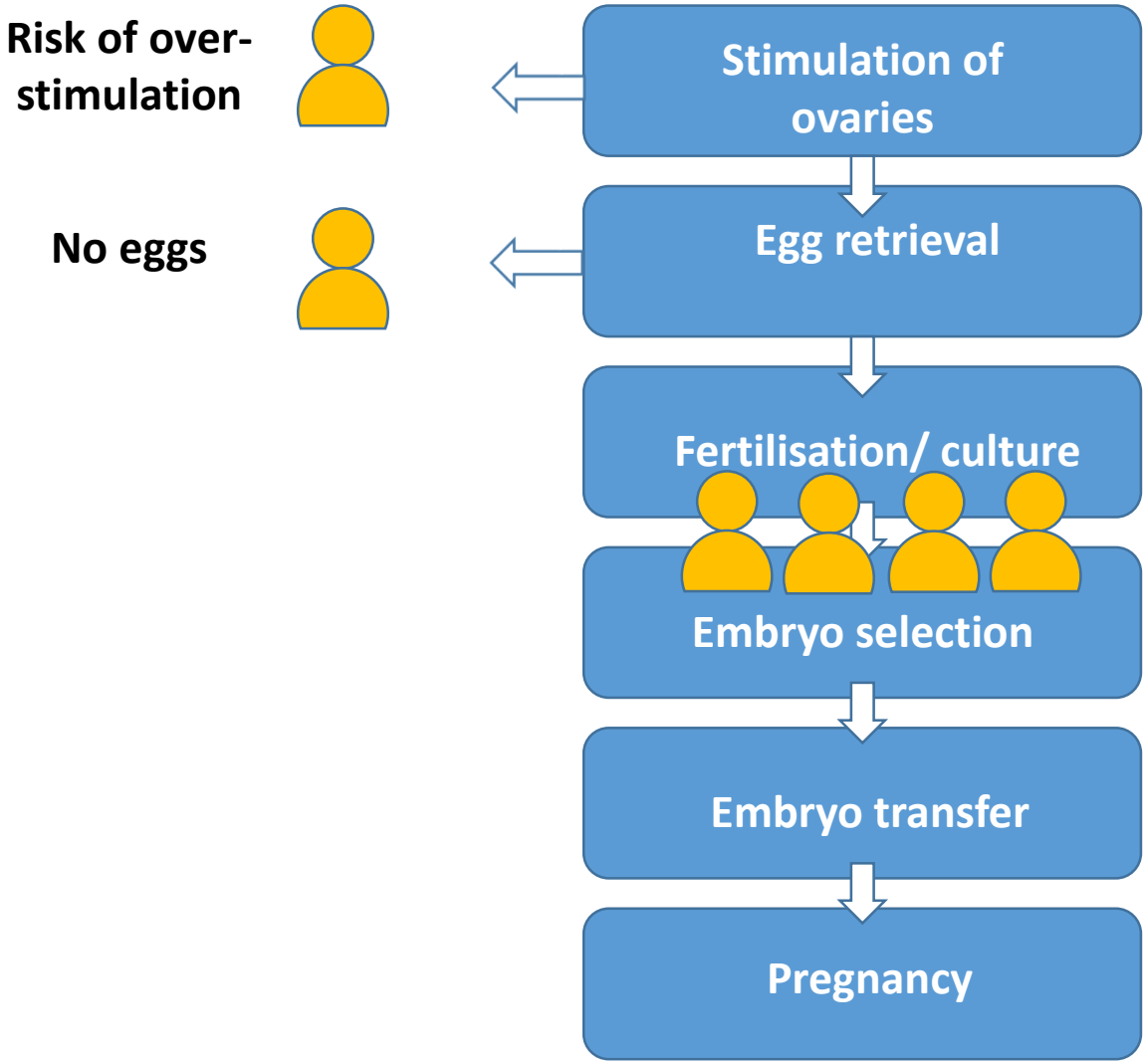
Fertilisation/ culture

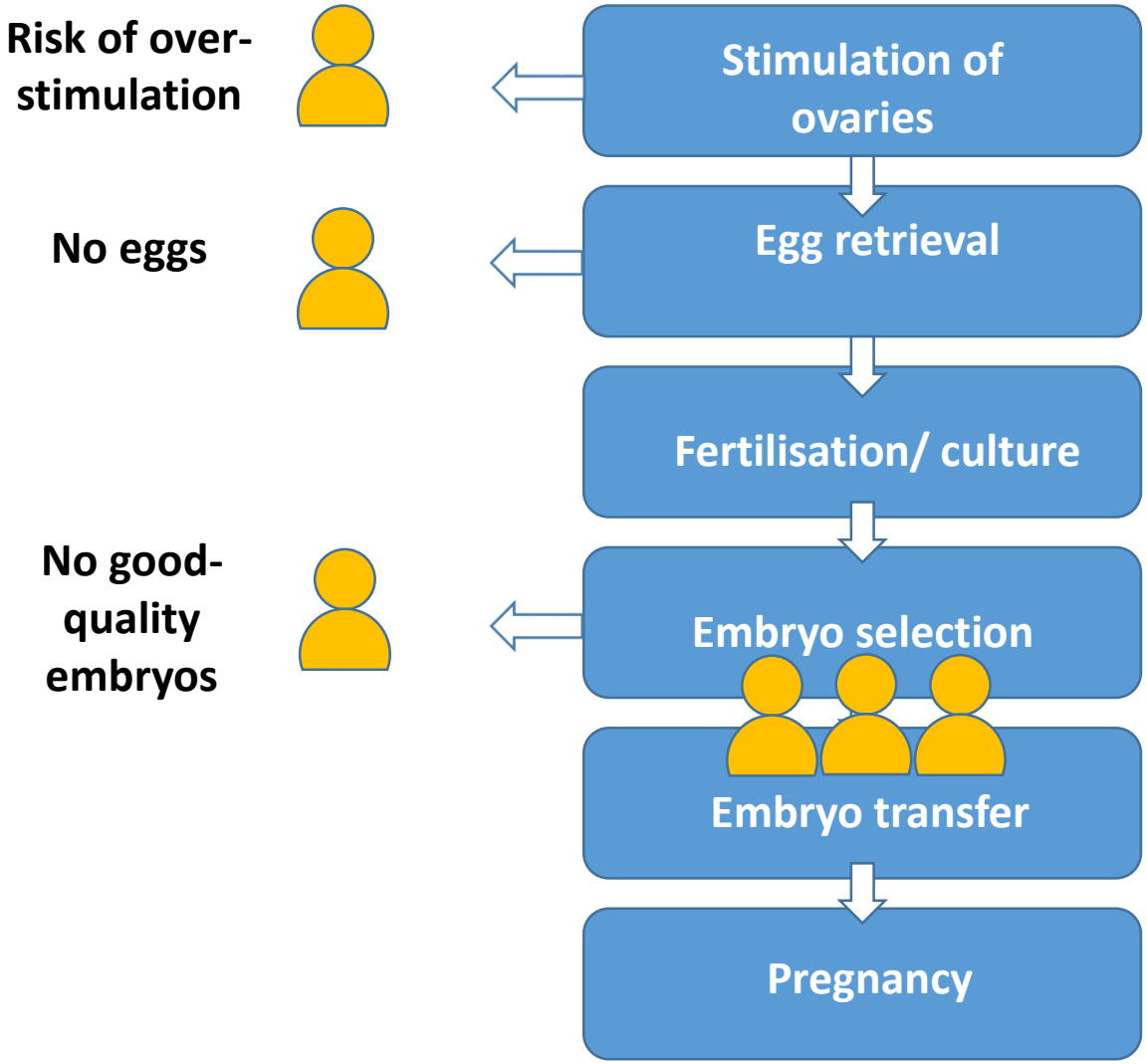
Embryo selection

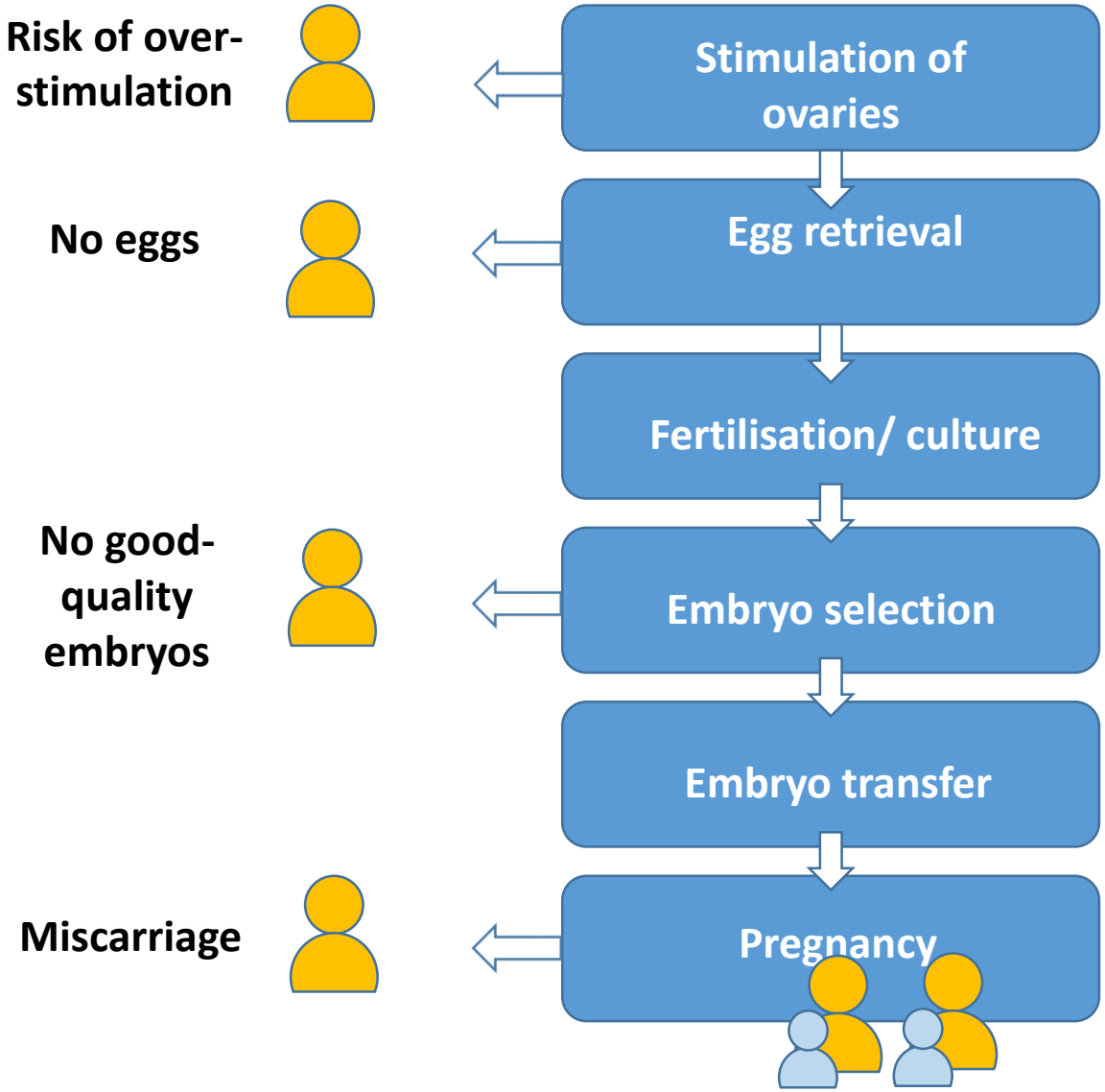
Embryo transfer

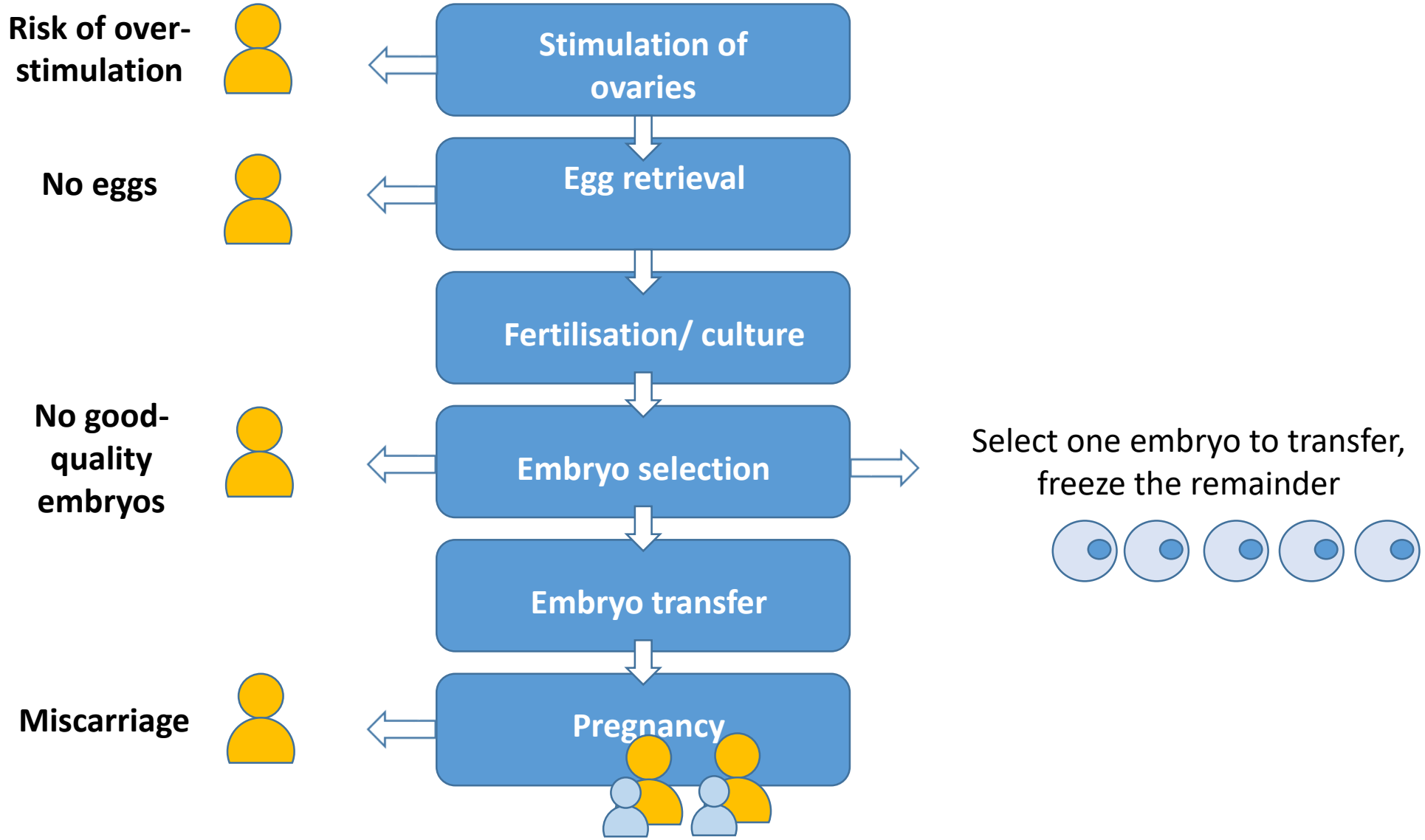
Pregnancy

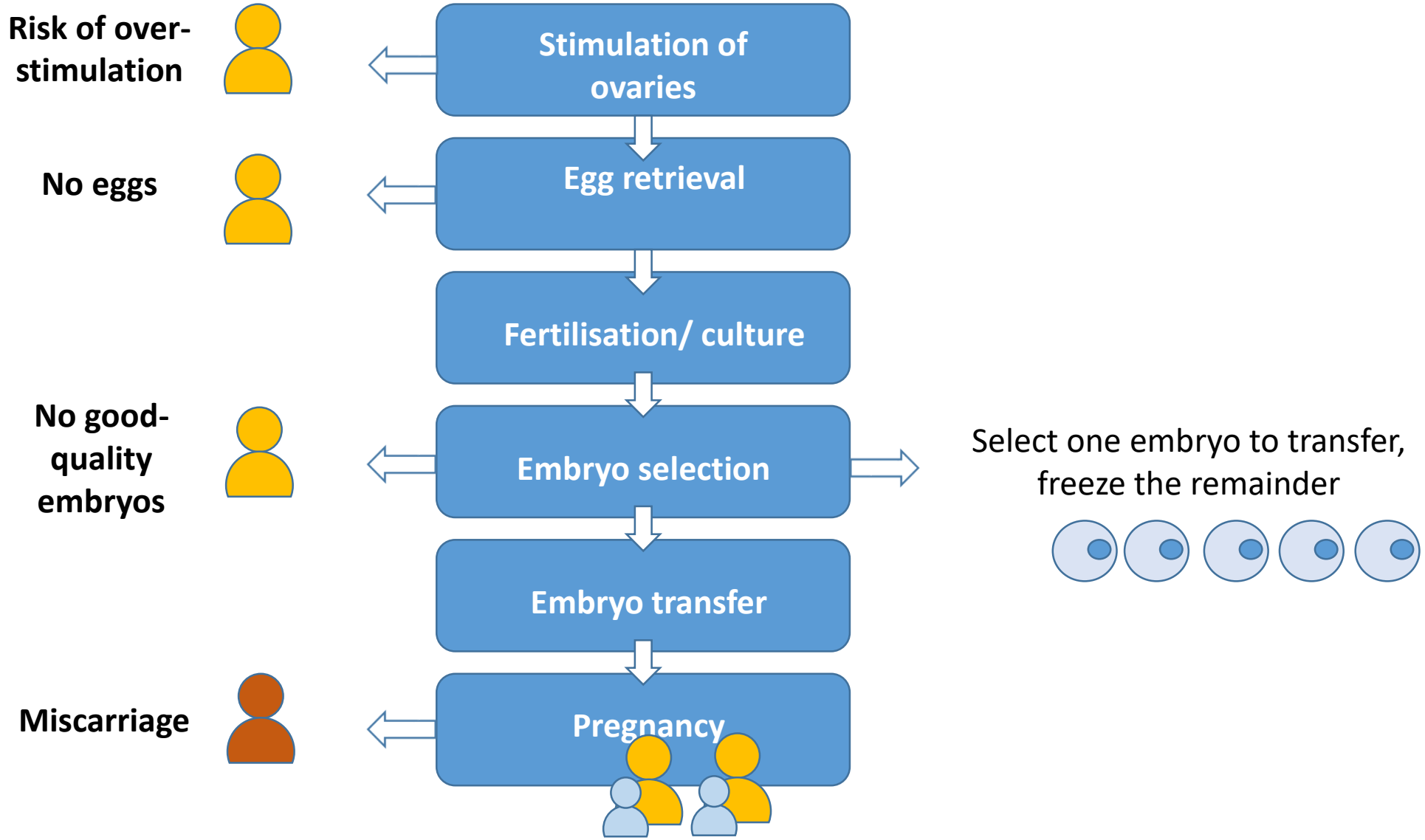

Risk of over-stimulation
Stimulation of ovaries
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy

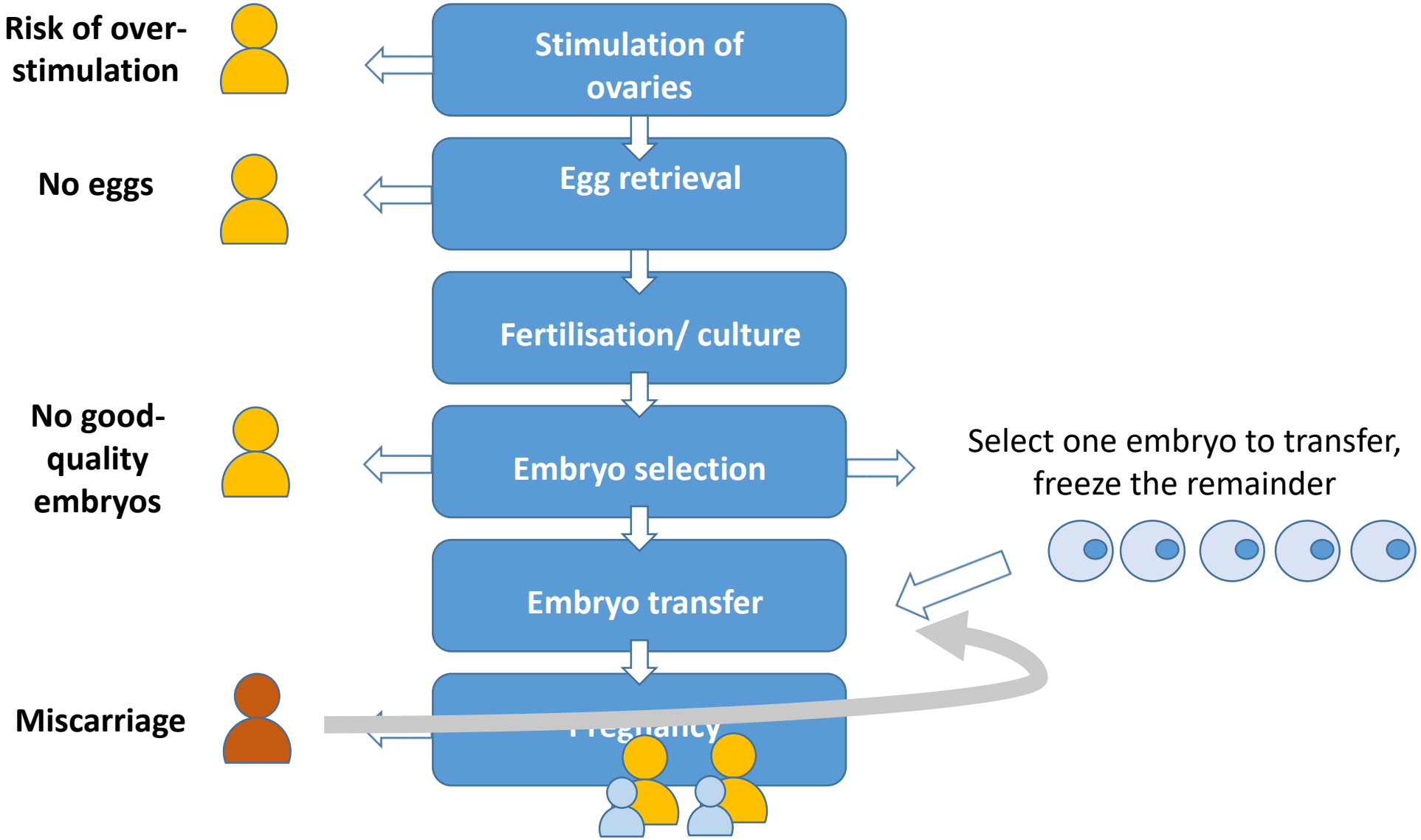

Risk of over-stimulation
No eggs
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy


Risk of over-stimulation
No eggs
No good-quality embryos
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy


Risk of over-stimulation
No eggs
No good-quality embryos
Miscarriage
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy


Risk of over-stimulation
No eggs
No good-quality embryos
Miscarriage
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
Select one embryo to transfer, freeze the remainder


Risk of over-stimulation
No eggs
No good-quality embryos
Miscarriage
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
Select one embryo to transfer, freeze the remainder


Risk of over-stimulation
Stimulation of ovaries
No eggs
Egg retrieval
Fertilisation/ culture
No good-quality embryos
Embryo selection
Select one embryo to transfer, freeze the remainder
Embryo transfer
Miscarriage
Pregnancy

Neonatal outcomes in IVF trials

human reproduction

ORIGINAL ARTICLE *Embryology*

Influence of embryo culture medium (G5 and HTF) on pregnancy and perinatal outcome after IVF: a multicenter RCT

Sander H.M. Kleijkers[1,†], Eleni Mantikou[2,†], Els Slappendel[3], Dimitri Consten[4], Jannie van Echten-Arends[5], Alex M. Wetzels[6], Madelon van Wely[2], Luc J.M. Smits[1], Aafke P.A. van Montfoort[1], Sjoerd Repping[2], John C.M. Dumoulin[1,†*], and Sebastiaan Mastenbroek[2,†*]

This trial looked at so-called **cumulative live birth (**after multiple embryo transfer attempts), but I'm going to ignore this aspect to streamline the discussion.

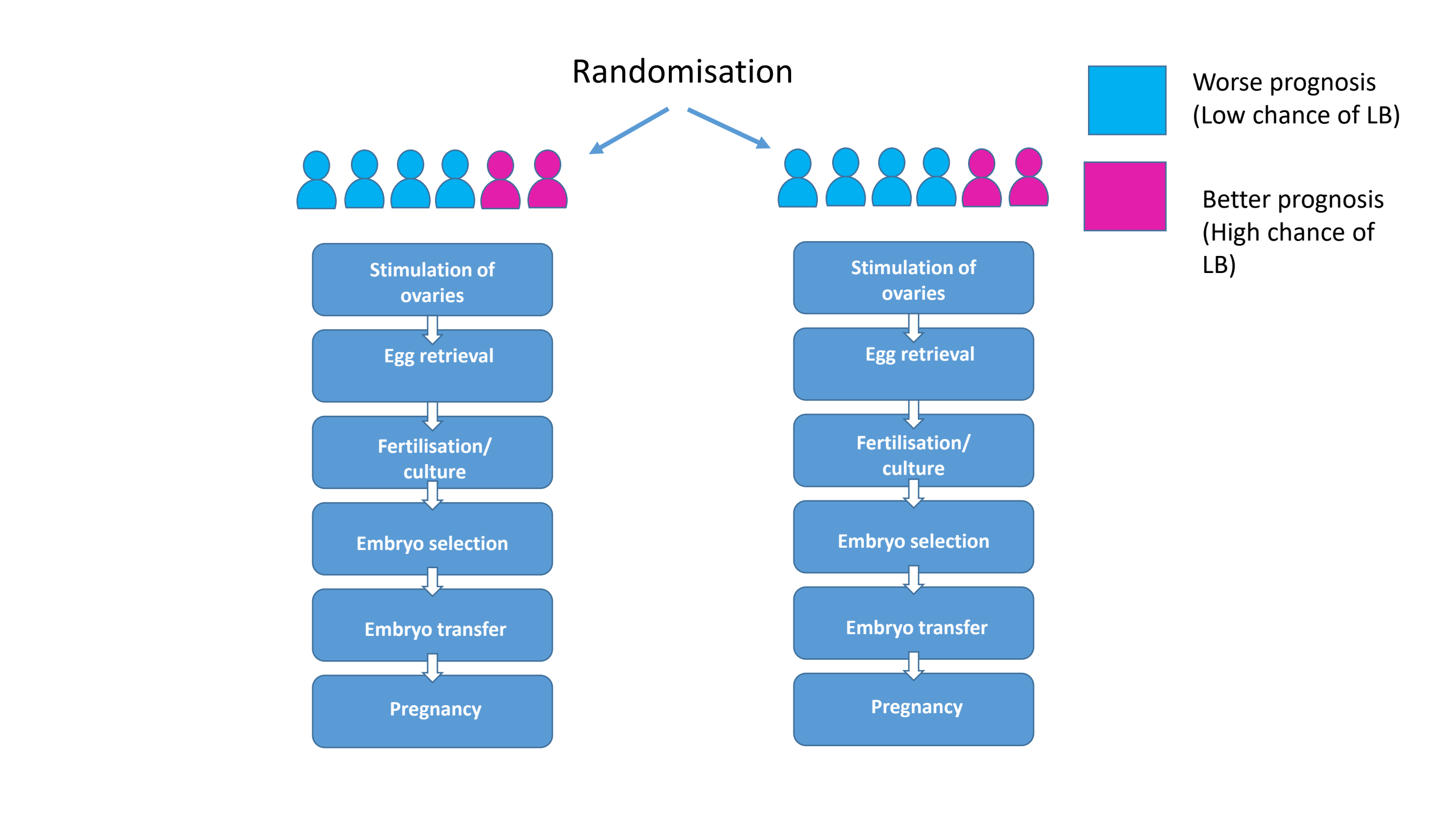
- Detailed info re: constituents are **not publically available**.

- Detailed info re: constituents are **not publically available**.

MAIN RESULTS AND THE ROLE OF CHANCE: The live birth rate was higher, albeit nonsignificantly, in couples assigned to G5 than in couples assigned to HTF (44.1% (184/417) versus 37.9% (159/419); RR: 1.2; 95% confidence interval (CI): 0.99–1.37; $P = 0.08$). Number of utilizable embryos per cycle (2.8 $\pm$ 2.3 versus 2.3 $\pm$ 1.8; $P < 0.001$), implantation rate after fresh embryo transfer (20.2 versus 15.3%; $P < 0.001$) and clinical pregnancy rate (47.7 versus 40.1%; RR: 1.2; 95% CI: 1.02–1.39; $P = 0.03$) were significantly higher for couples assigned to G5 compared with those assigned to HTF. Of the 383 live born children in this trial, birthweight data from 380 children (300 singletons (G5: 163, HTF: 137) and 80 twin children (G5: 38, HTF: 42)) were retrieved. Birthweight was significantly lower in the G5 group compared with the HTF group, with a mean difference of 158 g (95% CI: 42–275 g; $P = 0.008$). More singletons were born preterm in the G5 group (8.6% (14/163) versus 2.2% (3/137), but singleton birthweight adjusted for gestational age and gender (z-score) was also lower in the G5 than in the HTF group (-0.13 ± 0.08 versus 0.17 ± 0.08; $P = 0.008$).

- Suggests G5 increases live birth rate, but reduces birthweight (GEE to compare birthweight of infants)

- But is this an appropriate analysis?

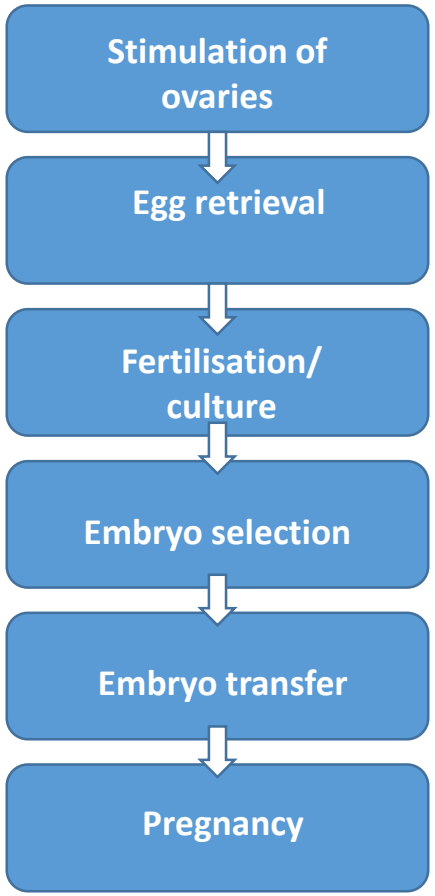

Randomisation
Worse prognosis
(Low chance of LB)
Better prognosis
(High chance of
LB)
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy

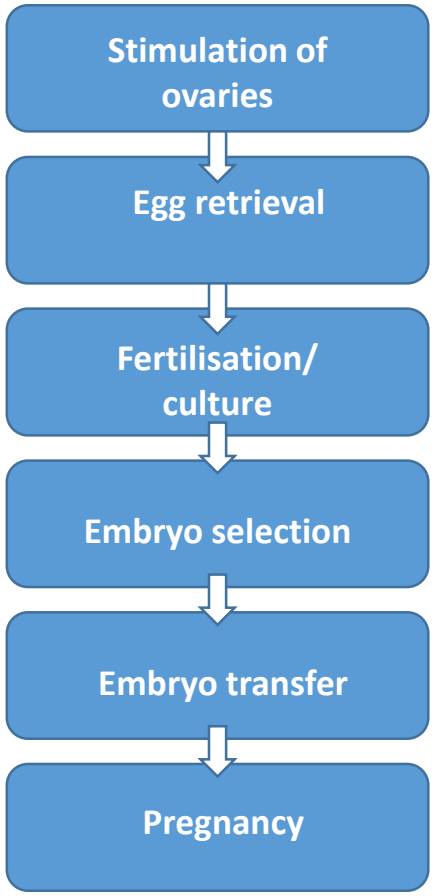
Randomisation

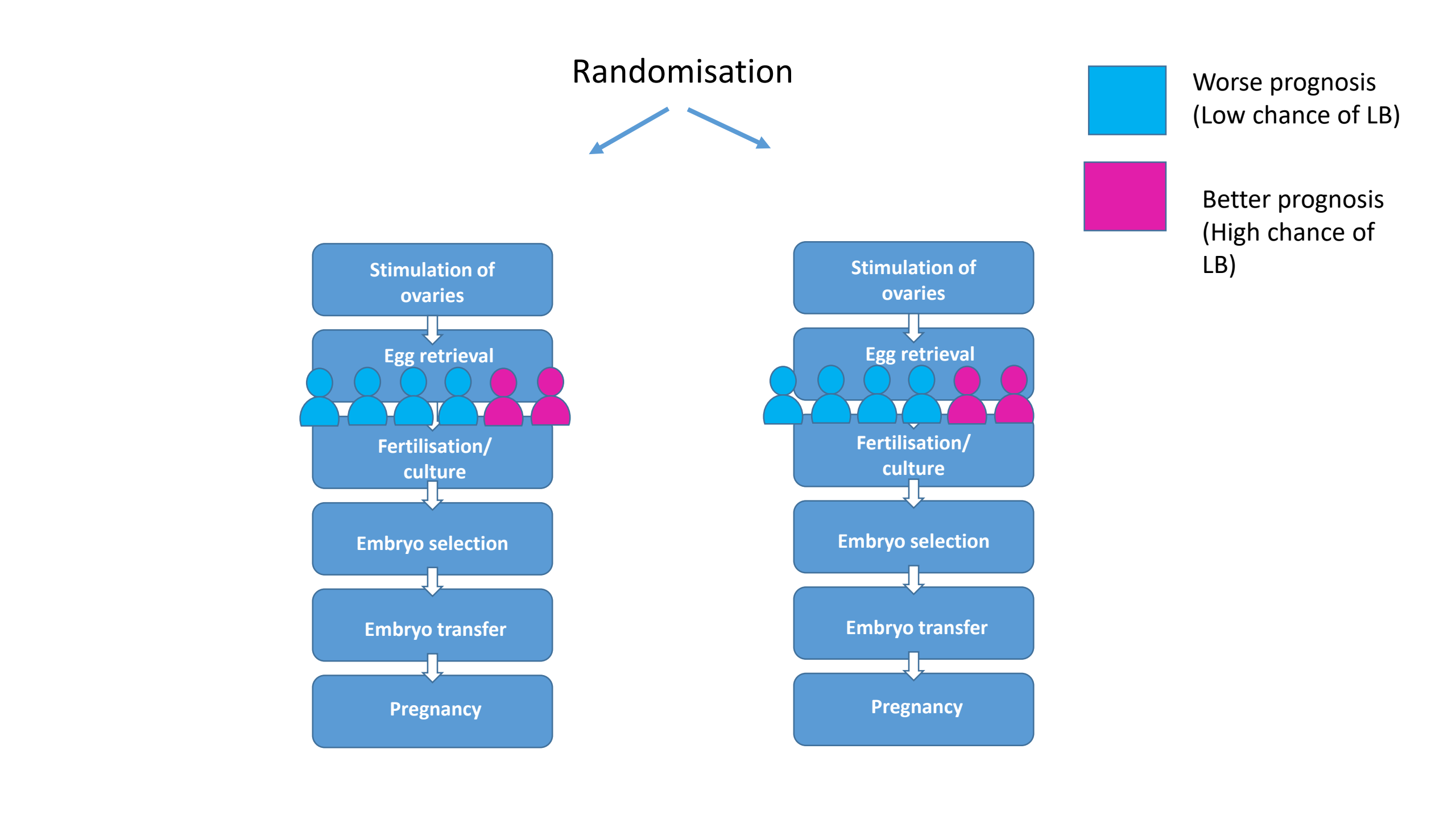
Any imbalances are random

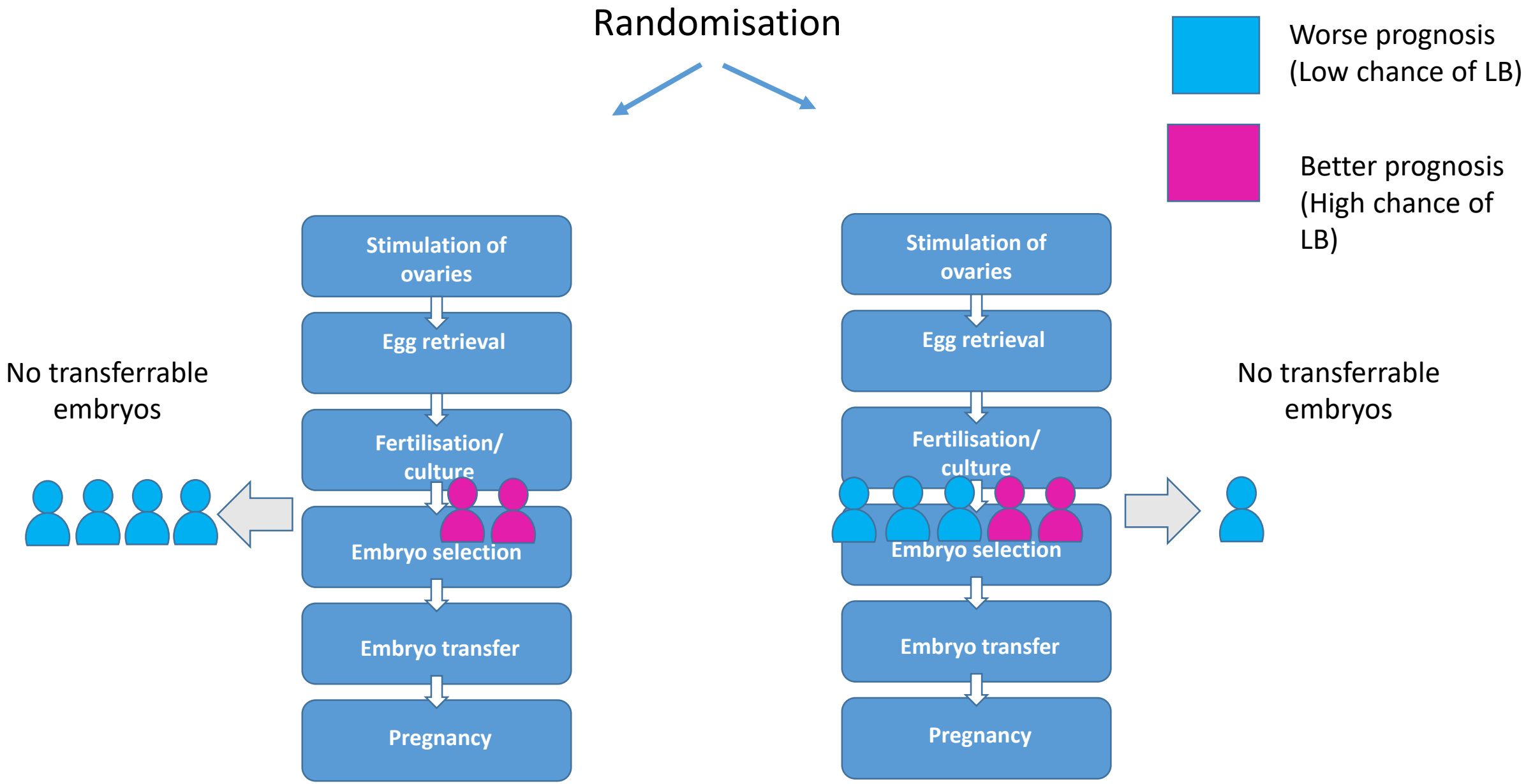
Worse prognosis (Low chance of LB)

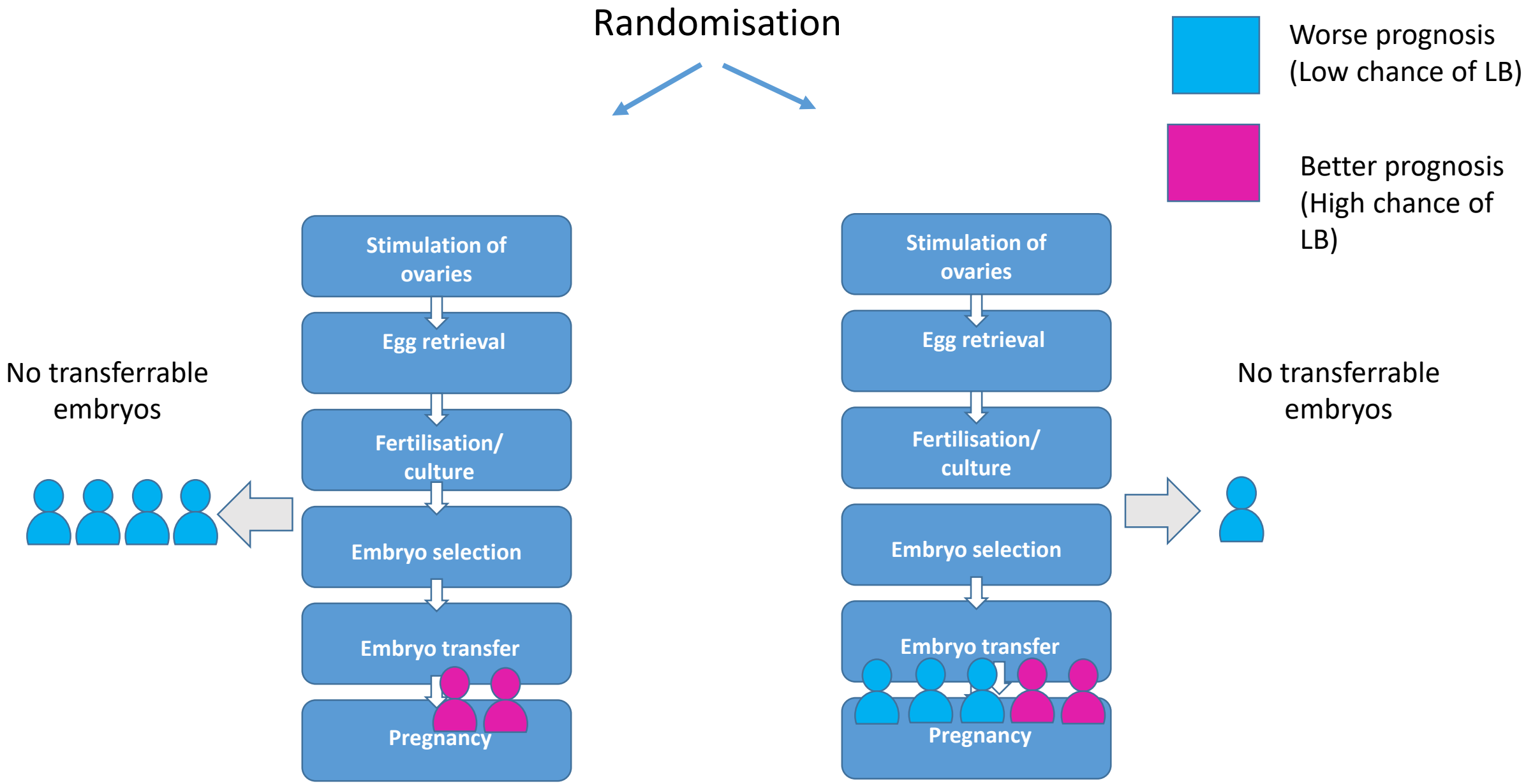
Better prognosis (High chance of LB)

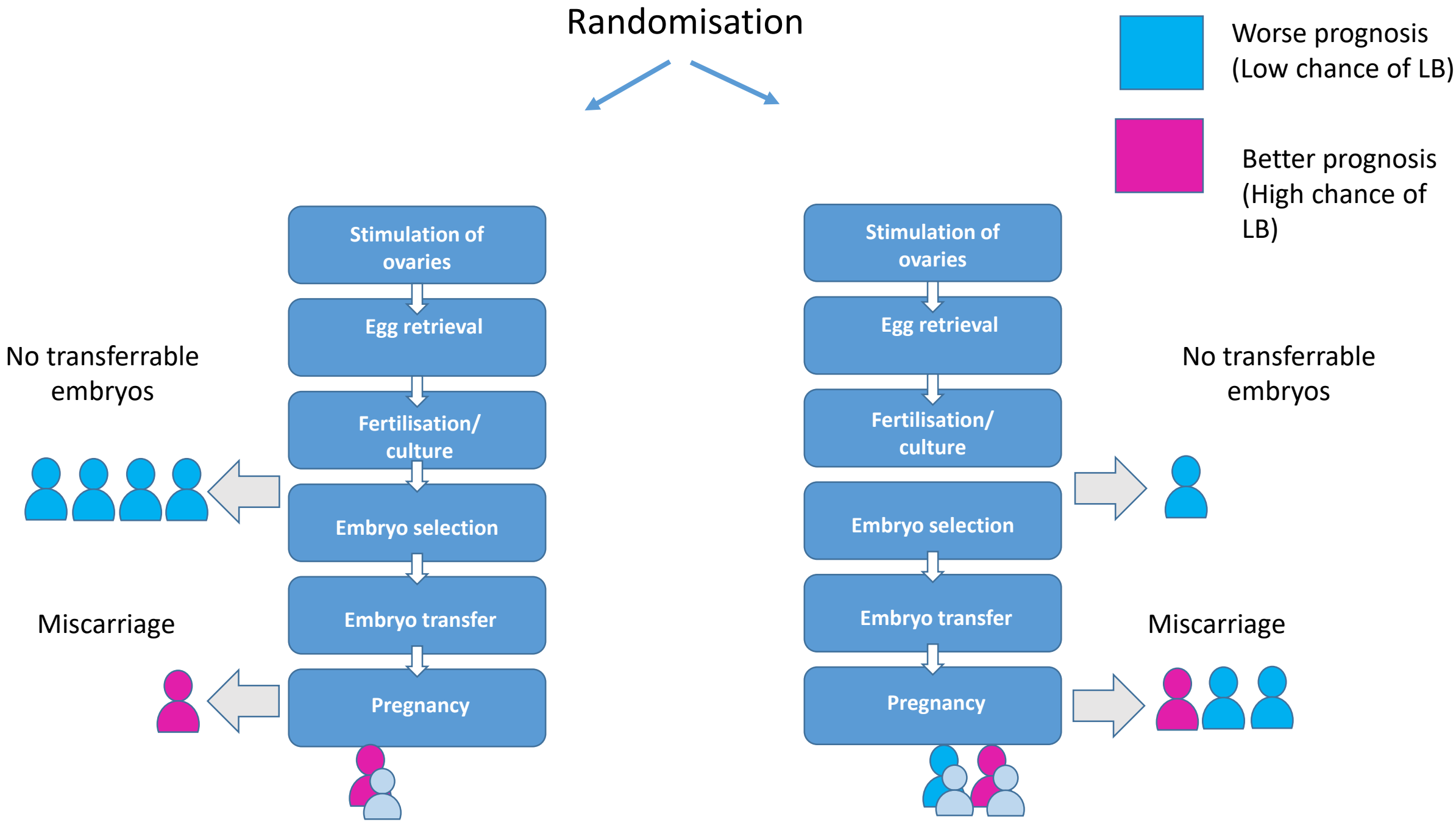


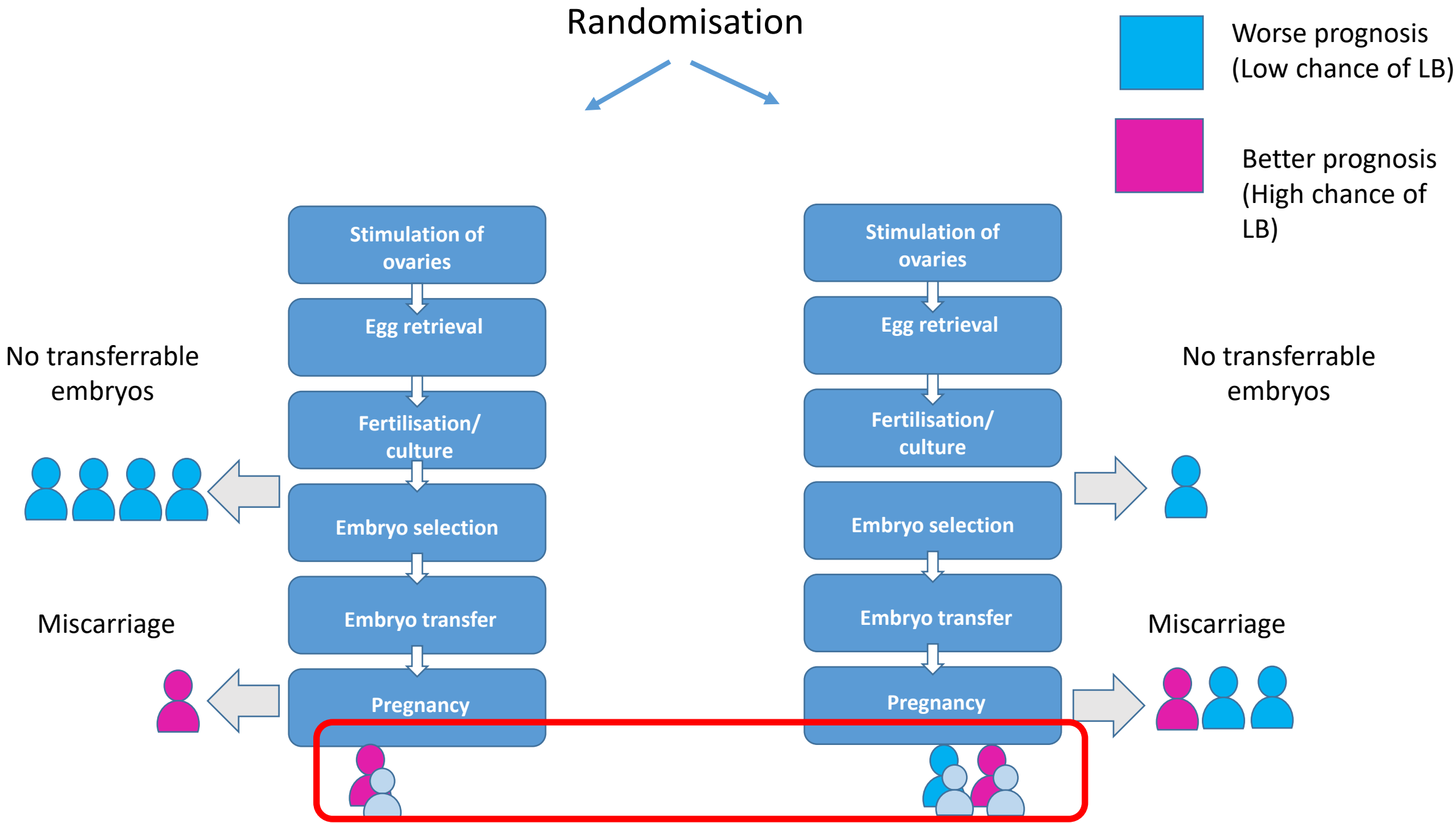


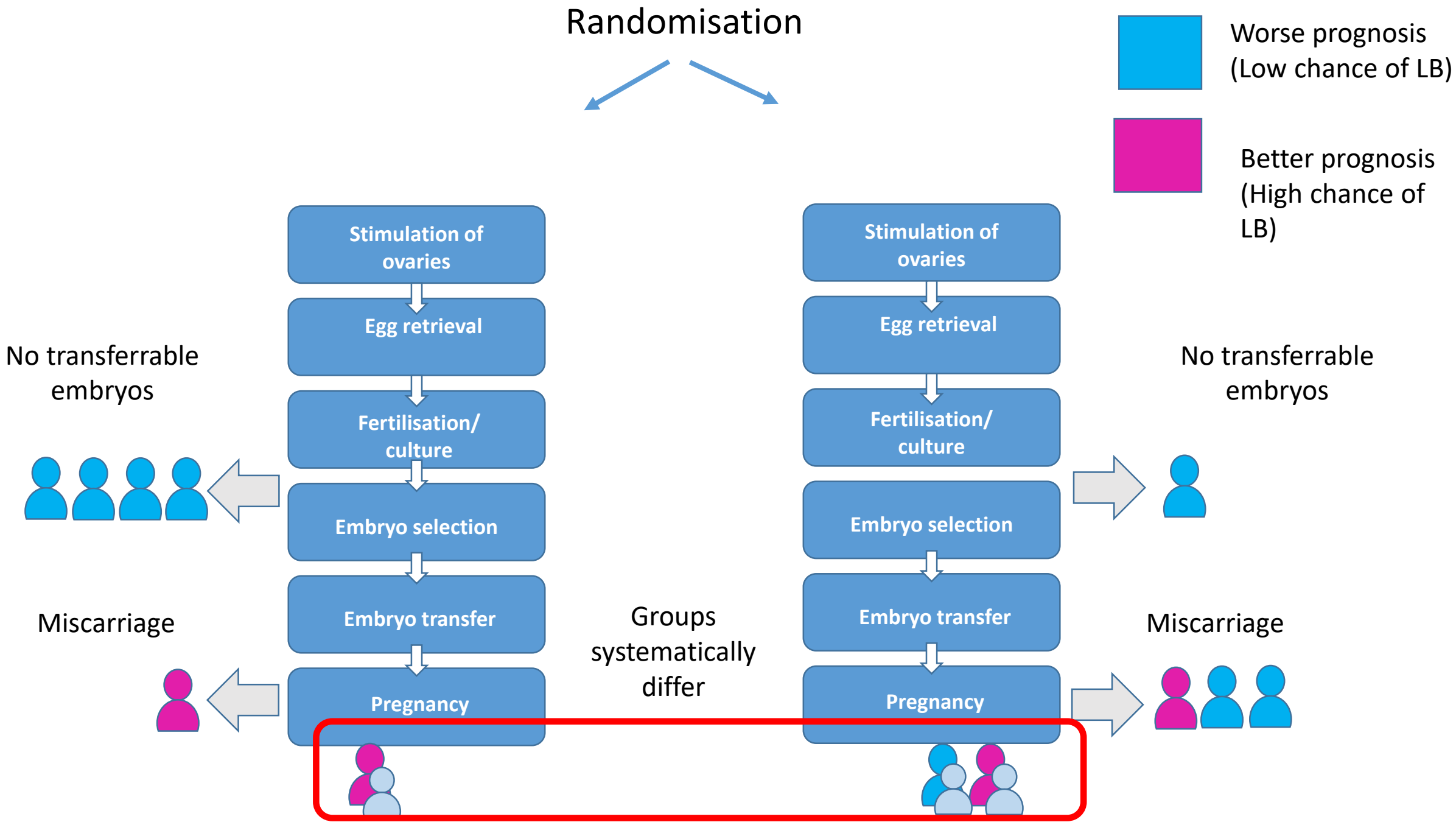

Randomisation
Worse prognosis
(Low chance of LB)
Better prognosis
(High chance of
LB)
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy

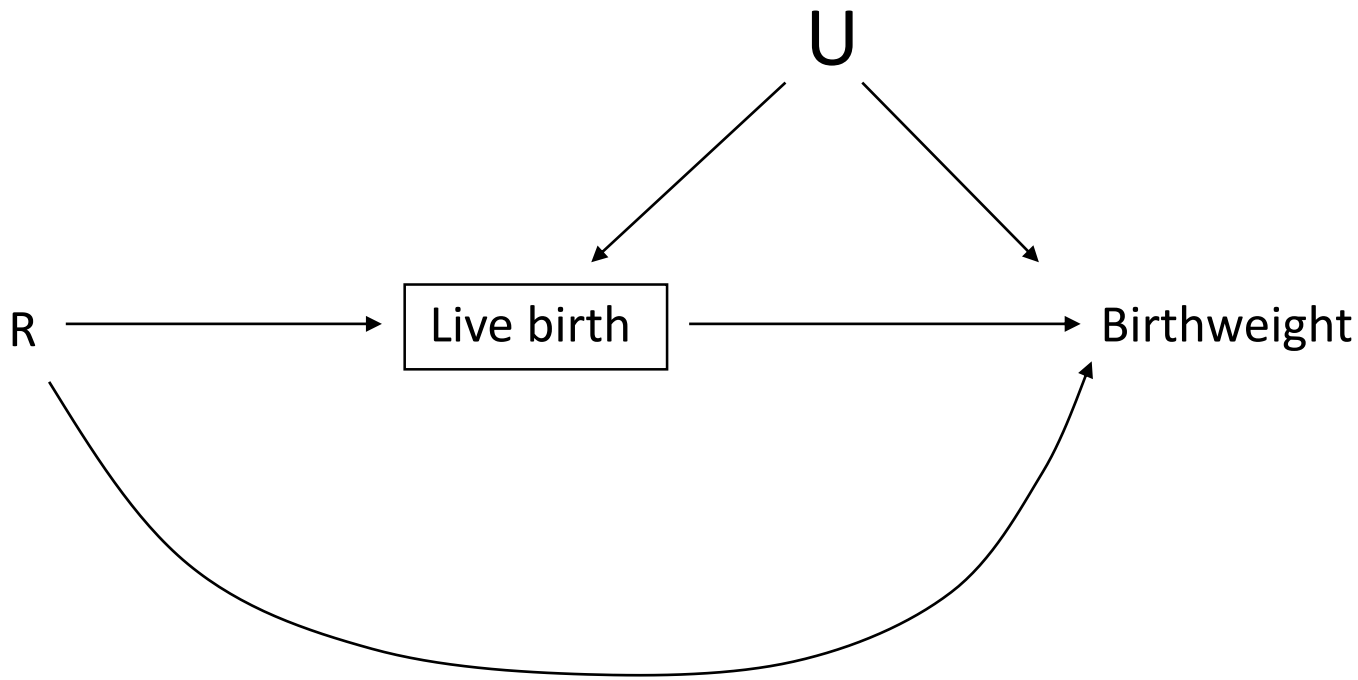

Randomisation
Worse prognosis
(Low chance of LB)
Better prognosis
(High chance of
LB)
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos

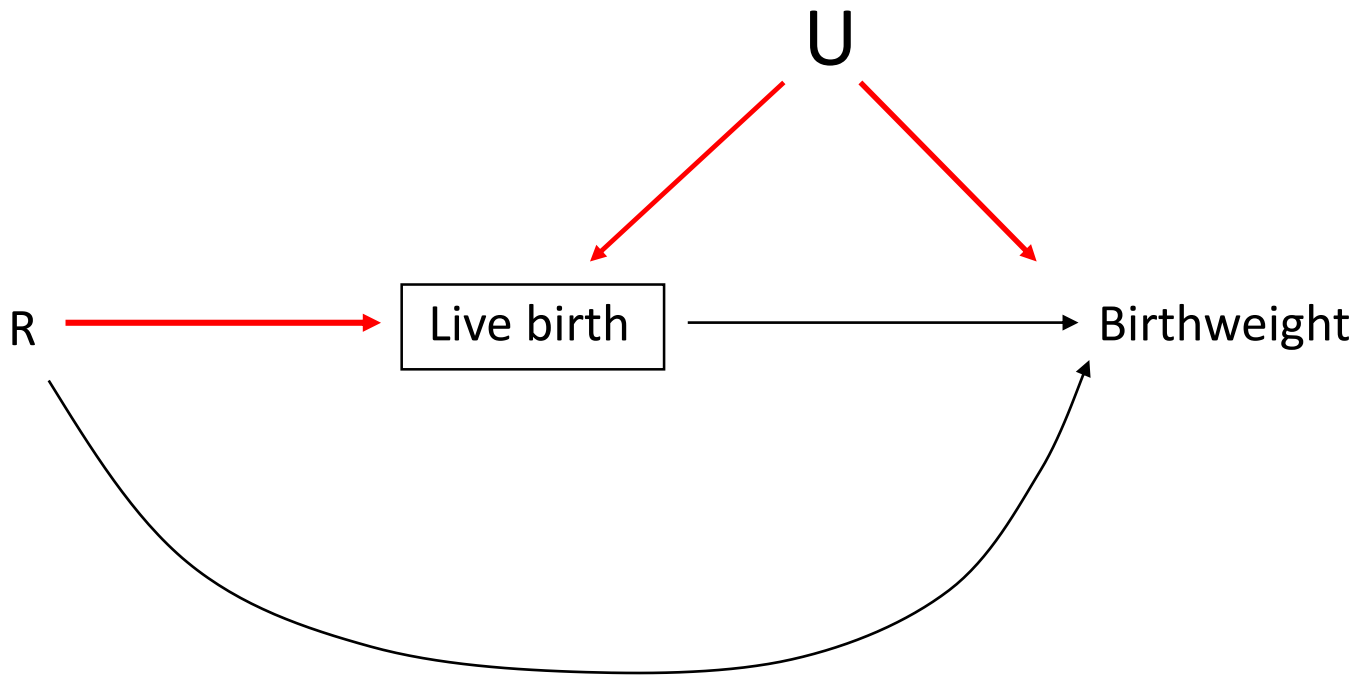

Randomisation
Worse prognosis
(Low chance of LB)
Better prognosis
(High chance of
LB)
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos


Randomisation
Worse prognosis
(Low chance of LB)
Better prognosis
(High chance of
LB)
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos
Miscarriage
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos
Miscarriage


Randomisation
Worse prognosis
(Low chance of LB)
Better prognosis
(High chance of
LB)
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos
Miscarriage
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos
Miscarriage


Randomisation
Worse prognosis
(Low chance of LB)
Better prognosis
(High chance of
LB)
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos
Miscarriage
Groups systematically differ
Stimulation of ovaries
Egg retrieval
Fertilisation/ culture
Embryo selection
Embryo transfer
Pregnancy
No transferrable embryos
Miscarriage


U
R
Live birth
Birthweight


U
R
Live birth
Birthweight

What do we want to estimate here?

- ICH E9(R1) addendum on estimands – framework for making our questions more precise.
- Declare strategy for handling **intercurrent events**: *post-randomisation events that affect interpretation or existence of outcome data* (Kahan, et al., 2024)

Five strategies for handling intercurrent events:

Strategy	Explanation
Treatment policy strategy	Estimate in all participants regardless of intercurrent events
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur
Composite strategy	Incorporate intercurrent event into outcome definition
While on treatment strategy	Treatment response before intercurrent event
Principal stratum strategy	Effect in people who would not have intercurrent event

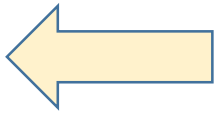
In this example, our intercurrent event would be ‘no live birth’

Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	
Composite strategy	Incorporate intercurrent event into outcome definition	
While on treatment strategy	Treatment response before intercurrent event	
Principal stratum strategy	Effect in people who would not have intercurrent event	

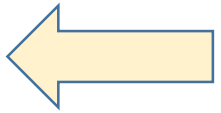
Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	
Composite strategy	Incorporate intercurrent event into outcome definition	
While on treatment strategy	Treatment response before intercurrent event	
Principal stratum strategy	Effect in people who would not have intercurrent event	



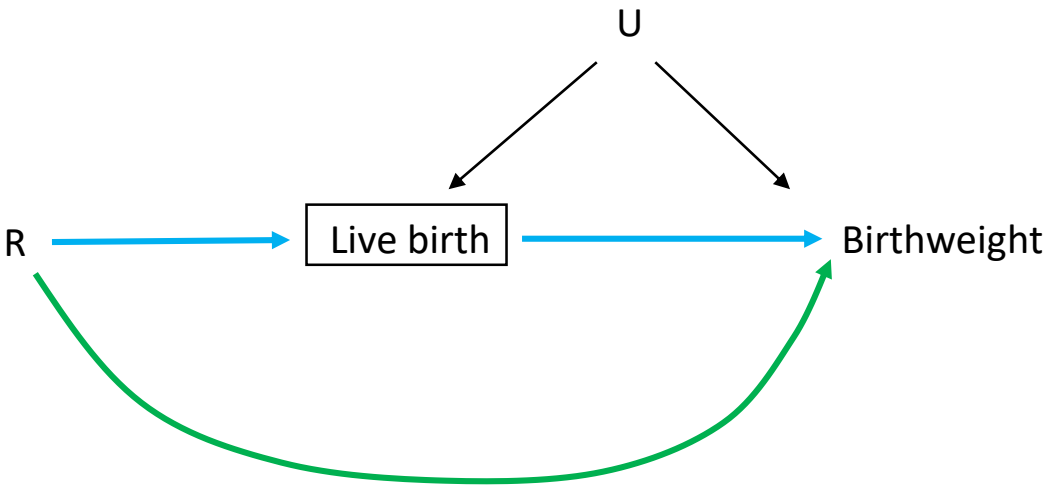
Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	Not useful? Suppose everyone has a baby from treatment (most do not)
Composite strategy	Incorporate intercurrent event into outcome definition	
While on treatment strategy	Treatment response before intercurrent event	
Principal stratum strategy	Effect in people who would not have intercurrent event	



The Effect of Prenatal treatments on offspring events in the presence of competing events: an application to a randomized trial of fertility therapies

Yu-Han Chiu[1,2,*], Mats J. Stensrud[1], Issa J. Dahabreh[1,3,4], Paolo Rinaudo[5], Michael P. Diamond[6], John Hsu[2,7], Sonia Hernández-Díaz[1], Miguel A. Hernán[1,8,9]

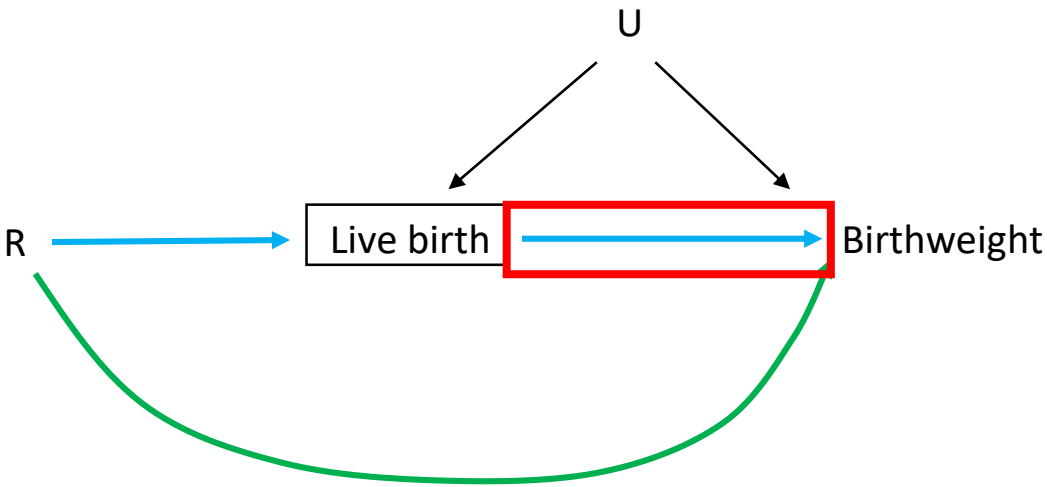


- "Separable direct and indirect effects" (Stensrud, et al., 2021)
- Idea: suppose treatment could be replaced by two hypothetical treatments, which exert effects via two pathways.
- Estimate using IPW.
- Challenges: is this hypothetical useful? And is live birth a mediator? (Snowden, et al., 2020)

Refs: doi: 10.1097/EDE.0000000000001222
doi: 10.1097/EDE.0000000000001223

The Effect of Prenatal treatments on offspring events in the presence of competing events: an application to a randomized trial of fertility therapies

Yu-Han Chiu[1,2,*], Mats J. Stensrud[1], Issa J. Dahabreh[1,3,4], Paolo Rinaudo[5], Michael P. Diamond[6], John Hsu[2,7], Sonia Hernández-Díaz[1], Miguel A. Hernán[1,8,9]

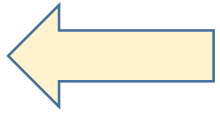


- "Separable direct and indirect effects" (Stensrud, et al., 2021)
- Idea: suppose treatment could be replaced by two hypothetical treatments, which exert effects via two pathways.
- Estimate using IPW.
- Challenges: is this hypothetical useful? And is live birth a mediator? (Snowden, et al., 2020)

Refs: doi: 10.1097/EDE.0000000000001222
doi: 10.1097/EDE.0000000000001223

Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	Not useful? Suppose everyone has a baby from treatment (most do not)
Composite strategy	Incorporate intercurrent event into outcome definition	How? 1) Assign an arbitrary birthweight if no birth? 2) dichotomise birthweight, outcome = live birth of healthy birthweight baby
While on treatment strategy	Treatment response before intercurrent event	
Principal stratum strategy	Effect in people who would not have intercurrent event	



What is the most relevant standard of success in assisted reproduction? The singleton, term gestation, live birth rate per cycle initiated: the BESST endpoint for assisted reproduction

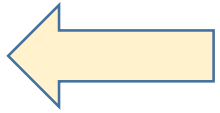
Jason K.Min[1], Sue A.Breheny[2], Vivien MacLachlan[2] and David L.Healy[3,4]

- Proposal from 2004: define outcome as singleton live birth of baby after 37 weeks of gestation
- Objections: is a healthy baby born before 37 weeks really a failure? (Griesinger, et al., 2004)
- Interpretation: Intervention could increase live birth but be worse on the composite
- Unit of analysis issue in Kleijkers – live birth (event) measured at the level of a treated individual, healthy birthweight at the level of the infant (twins possible) → redefine as e.g. "live birth event with all babies healthy birthweight"

Ref: doi.org/10.1093/humrep/deh237

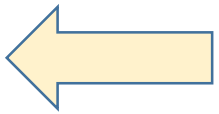
Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	Not useful? Suppose everyone has a baby from treatment (most do not)
Composite strategy	Incorporate intercurrent event into outcome definition	How? 1) Assign an arbitrary birthweight if no birth? 2) dichotomise birthweight, outcome = live birth of healthy birthweight baby
While on treatment strategy	Treatment response before intercurrent event	
Principal stratum strategy	Effect in people who would not have intercurrent event	



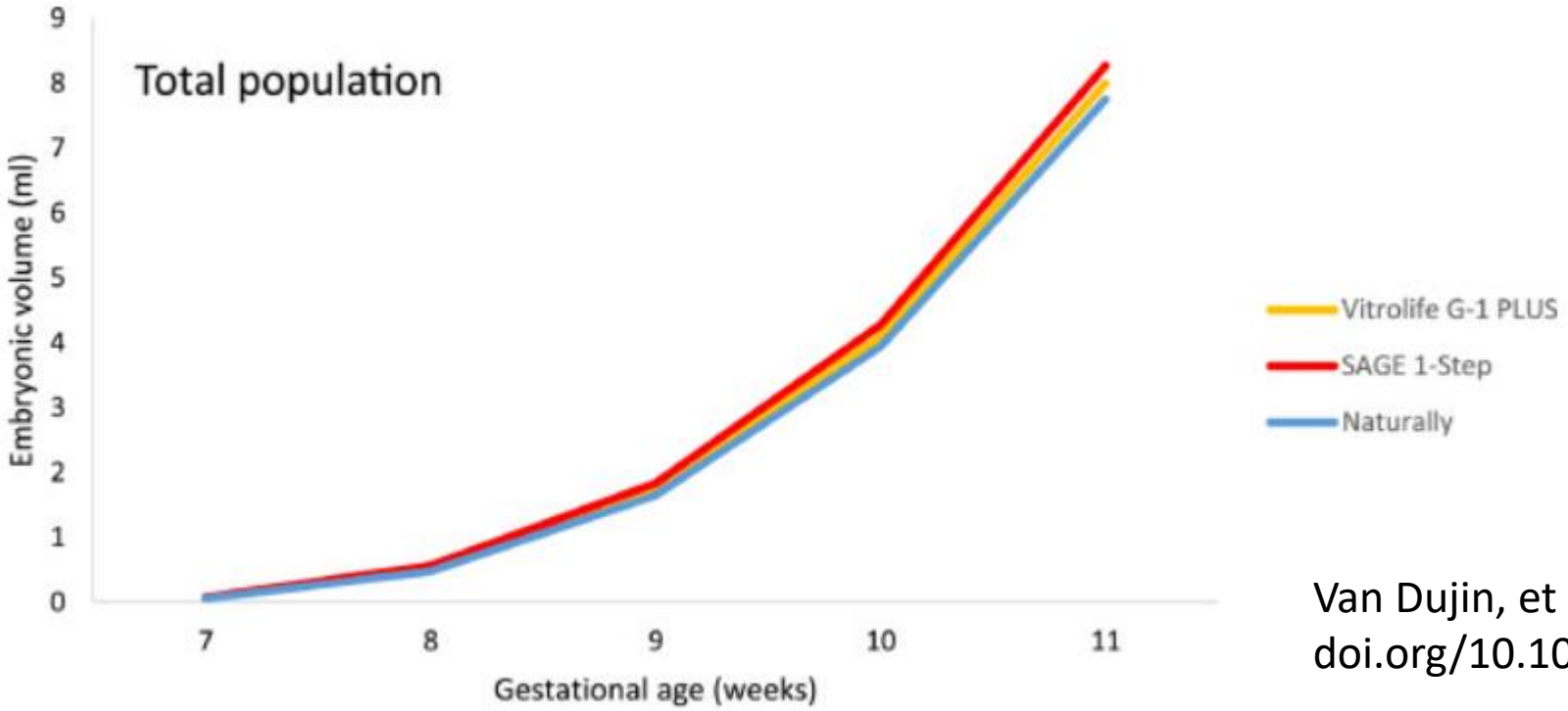
Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	Not useful? Suppose everyone has a baby from treatment (most do not)
Composite strategy	Incorporate intercurrent event into outcome definition	How? 1) Assign an arbitrary birthweight if no birth? 2) dichotomise birthweight, outcome = live birth of healthy birthweight baby
While on treatment strategy	Treatment response before intercurrent event	There is no “birthweight” prior to live birth, but…
Principal stratum strategy	Effect in people who would not have intercurrent event	



IVF culture medium affects human intrauterine growth as early as the second trimester of pregnancy†

Ewka C.M. Nelissen[1,*], Aafke P.A. Van Montfoort[1], Luc J.M. Smits[2], Paul P.C.A. Menheere[3], Johannes L.H. Evers[1], Edith Coonen[4], Josien G. Derhaag[1], Louis L. Peeters[1], Audrey B. Coumans[1], and John C.M. Dumoulin[1]

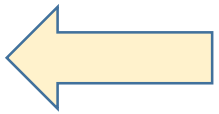


- Can use ultrasound to look at development of embryo or foetus in utero.
- Effects of culture media on development can be detected before live birth.
- Might allow for some sort of while-on-treatment analysis.
- However, many participants will have outcome truncated prior to this.
- In Kleijkers 2016, 44% and 50% had no pregnancy at all.

Van Dujin, et al. 2022
doi.org/10.1016/j.rbmo.2022.06.003

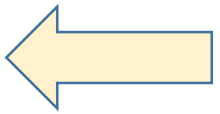
Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	Not useful? Suppose everyone has a baby from treatment (most do not)
Composite strategy	Incorporate intercurrent event into outcome definition	How? 1) Assign an arbitrary birthweight if no birth? 2) dichotomise birthweight, outcome = live birth of healthy birthweight baby
While on treatment strategy	Treatment response before intercurrent event	There is no “birthweight” prior to live birth, but…
Principal stratum strategy	Effect in people who would not have intercurrent event	



Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	Not useful? Suppose everyone has a baby from treatment (most do not)
Composite strategy	Incorporate intercurrent event into outcome definition	How? 1) Assign an arbitrary birthweight if no birth? 2) dichotomise birthweight, outcome = live birth of healthy birthweight baby
While on treatment strategy	Treatment response before intercurrent event	There is no “birthweight” prior to live birth, but…
Principal stratum strategy	Effect in people who would not have intercurrent event	Survivor average causal effect (SACE)



Survivor average causal effect (SACE)

- Effect in infants who would be born **regardless of allocation** (e.g. Tchetgen Tchetgen 2014).
- Estimation requires strong assumptions e.g. **montonicity** – anyone having a baby with HTF would also have a baby with G5.
- But regardless, is this really a quantity of interest anyway?
- If we are introducing a new fertility treatment, we usually hope it increases the number of births.
- So we are also interested in outcomes in infants born as a result of new interventions, who otherwise would not have been born (Kahan, personal correspondence).
- **How to estimate that?**

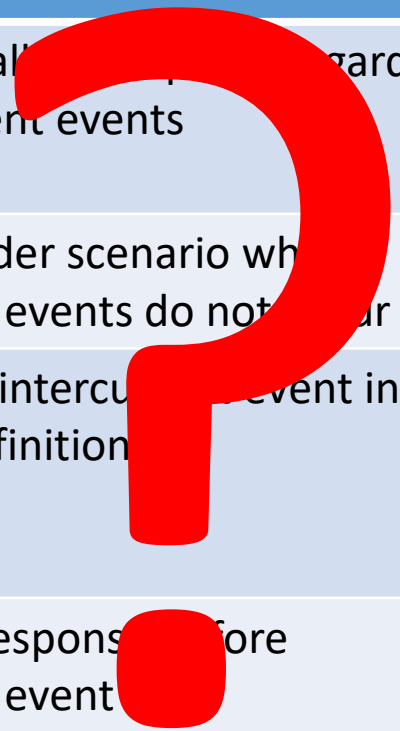
Ref: doi: 10.1002/sim.6181

Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all participants regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	Not useful? Suppose everyone has a baby from treatment (most do not)
Composite strategy	Incorporate intercurrent event into outcome definition	How? 1) Assign an arbitrary birthweight if no birth? 2) dichotomise birthweight, outcome = live birth of healthy birthweight baby
While on treatment strategy	Treatment response before intercurrent event	There is no “birthweight” prior to live birth, but…
Principal stratum strategy	Effect in people who would not have intercurrent event	Survivor average causal effect (SACE)

Application to culture media trial

Strategy	Explanation	In this context
Treatment policy strategy	Estimate in all [illegible] regardless of intercurrent events	Not possible in this context? Outcome only defined for people who have births
Hypothetical strategy	Estimate under scenario where intercurrent events do not occur	Not useful? Suppose everyone has a baby from treatment (most do not)
Composite strategy	Incorporate intercurrent event into outcome definition	How? 1) Assign an arbitrary birthweight if no birth? 2) dichotomise birthweight, outcome = live birth of healthy birthweight baby
While on treatment strategy	Treatment response before intercurrent event	There is no “birthweight” prior to live birth, but…
Principal stratum strategy	Effect in people who would not have intercurrent event	Survivor average causal effect (SACE)



- Viable intrauterine pregnancy confirmed by ultrasound. Reporting singleton pregnancy, twin pregnancy, and higher multiple pregnancy.
- Pregnancy loss. Reporting ectopic pregnancy, miscarriage, stillbirth, and termination of pregnancy
- Live birth.
- Gestational age at delivery.
- Birthweight.
- Neonatal mortality.
- Major congenital anomaly.

* When applicable → time to pregnancy leading to live birth.

Core outcome set for infertility trials: https://doi.org/10.1016/j.fertnstert.2020.11.013

- Viable intrauterine pregnancy confirmed by ultrasound. Reporting singleton pregnancy, twin pregnancy, and higher multiple pregnancy.
- Pregnancy loss. Reporting ectopic pregnancy, miscarriage, stillbirth, and termination of pregnancy
- Live birth.
- Gestational age at delivery.
- Birthweight.
- Neonatal mortality.
- Major congenital anomaly.

* When applicable → time to pregnancy leading to live birth.

Core outcome set for infertility trials: https://doi.org/10.1016/j.fertnstert.2020.11.013

What to advise?

Estimands framework: conceptually clear, useful causal inference appears to be very difficult here.

Proposal 1: Descriptive approach

- Abandon efforts to study treatment effects
- Summarise neonatal outcomes by arm ('is there anything alarming going on'?)
- Interpret with reference to participant characteristics and effects on pregnancy, live birth
- Requires us to abandon logic of experiments

Proposal 2: Composite approach

- Define composite (e.g. live birth of healthy birthweight infants)
- Interpret result with reference to effect on live birth (regardless of birthweight)
- Familiar logic – e.g. miscarriage per woman randomised could be reduced by reducing pregnancies
- Ordinal version of this? No LB->LB wo healthy BW -> LB with healthy bw

Closing remarks

- Estimands framework useful for considering options – despite not offering a clear solution in this case.

- Various other challenges may benefit from an estimands perspective (handling of natural pregnancies, timing of randomisation, analysis of time to event across multiple treatment events). Worth exploring further.

- I'd be interested to hear opinions on the example discussed here.

jack.wilkinson@manchester.ac.uk or @jd_wilko