INTERFACE

royalsocietypublishing.org/journal/rsif

Research

Cite this article: Leahy BD, Racowsky C, Needleman D. 2021 Inferring simple but precise quantitative models of human oocyte and early embryo development. *J. R. Soc. Interface* **18**: 20210475.
https://doi.org/10.1098/rsif.2021.0475

Received: 8 June 2021
Accepted: 16 August 2021

Subject Category:
Life Sciences–Physics interface

Subject Areas:
biophysics, biomathematics

Keywords:
pre-implantation development, human embryos, *in-vitro* fertilization, quantitative biology

Author for correspondence:
Brian D. Leahy
e-mail: b.d.leahy0@gmail.com

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5577702.

THE ROYAL SOCIETY
PUBLISHING

Inferring simple but precise quantitative models of human oocyte and early embryo development

Brian D. Leahy[1,2], Catherine Racowsky[3,4] and Daniel Needleman[1,2,5]

[1]Department of Molecular and Cellular Biology, and [2]SEAS, Harvard University, Cambridge, MA, USA
[3]Brigham Women's Hospital, Boston, MA, USA
[4]Harvard Medical School, Boston, MA, USA
[5]Center for Computational Biology, Flatiron Institute, New York, NY, USA

BDL, 0000-0002-0070-0985

Macroscopic, phenomenological models are useful as concise framings of our understandings in fields from statistical physics to finance to biology. Constructing a phenomenological model for development would provide a framework for understanding the complicated, regulatory nature of oogenesis and embryogenesis. Here, we use a data-driven approach to infer quantitative, precise models of human oocyte maturation and pre-implantation embryo development, by analysing clinical *in-vitro* fertilization (IVF) data on 7399 IVF cycles resulting in 57 827 embryos. Surprisingly, we find that both oocyte maturation and early embryo development are quantitatively described by simple models with minimal interactions. This simplicity suggests that oogenesis and embryogenesis are composed of modular processes that are relatively siloed from one another. In particular, our analysis provides strong evidence that (i) pre-antral follicles produce anti-Müllerian hormone independently of effects from other follicles, (ii) oocytes mature to metaphase-II independently of the woman's age, her BMI and other factors, (iii) early embryo development is memoryless for the variables assessed here, in that the probability of an embryo transitioning from its current developmental stage to the next is independent of its previous stage. Our results both provide insight into the fundamentals of oogenesis and embryogenesis and have implications for the clinical IVF.

1. Introduction

Understanding the manner by which a multicellular organism develops from a single cell is one of the grand challenges of biology. In mammals, this process begins with oogenesis inside the female, which results in an egg that becomes an embryo after fertilization. Early embryo development in mammals, including humans, is self-organized [1,2]: the course of events that unfold are governed by the embryo's internal dynamics and can proceed without external signals. Oogenesis and early embryogenesis have been studied from diverse perspectives, including molecular genetic, cell biological, chemical and mechanical [3–13]. Despite the vast amount of knowledge that has been obtained, many basic questions remain, including: what determines which oocytes are selected for ovulation? How is the timing of embryonic events regulated? How are oogenesis and embryogenesis negatively impacted by age and disease? Answering these will provide fundamental insight and have strong implications for evolution and for medical treatments of infertility. However, these issues are difficult to study using the molecular approaches that are the mainstay of current research, because of the integrated nature of the problems they pose concerning the overall trajectory of development.

An alternative to the microscopic, molecular perspective is to develop a macroscopic, phenomenological understanding. Such an approach has been productive in diverse areas from statistical physics [14] to finance [15] to some fields of biology [16–18], including protein evolution [19] and cell-size control in bacteria [20–22]. These phenomenological models focus on describing a few key variables that subsume the detailed descriptions of the component parts, e.g. temperature rather than the motions of individual molecules, or market volatility rather than the financial decisions of individual companies or investors. One significant concern is that the great complexity of oogenesis and embryogenesis might make simple, phenomenological descriptions inapplicable. Furthermore, the validity of the phenomenological approach can only be determined by developing models and rigorously testing them. This requires a large amount of quantitative data, which is difficult to obtain from oocytes and embryos in model organisms.

Here, we overcome this challenge by leveraging a large dataset from 7399 routine clinical *in-vitro* fertilization (IVF) treatment cycles resulting in 98 264 oocytes and 57 827 embryos; the treatments were performed from 2012 to 2017 at the Brigham and Women's Hospital in Boston, MA. We show that the data can be quantitatively described using simple, phenomenological models. The large variance in the data allows testing these models over a wide range of physiological conditions. The models we develop are Bayesian networks, a form of probabilistic graphical models which represents conditional dependencies by a directed, acyclic graph. We infer models directly from the data, making little use of prior knowledge. The resulting models recapitulate well-established aspects of oocyte and embryo development. Moreover, the resultant models are sparse: only one or two factors directly impact physiological processes. This implies that human oogenesis and embryogenesis are highly modular. Our analysis leads to a number of additional, surprising conclusions. We present strong evidence that:

(i) Each pre-antral follicle produces anti-Müllerian hormone (AMH) independently of effects from other follicles. This argues that AMH is a faithful indicator of the number of pre-antral follicles, consistent with its physiological role in regulating follicle recruitment and supporting its clinical use as a measure of ovarian reserve [23–26].

(ii) While the number of oocytes released from follicles depends on many factors, the probability that a released oocyte matures to metaphase-II is independent of the patient's age, BMI and other, external factors. This argues that physiological processes that are correlated with these external factors, such as mitochondrial metabolism and aneuploidy [27–29], do not significantly impact meiosis resumption.

(iii) After oocytes are fertilized, the probability of successfully transitioning from one embryonic developmental stage to the next depends on the embryo's present state, but not on its earlier state. Thus, embryo development is memoryless, at least for the variables examined here. This argues that clinical embryo selection procedures need only consider the state of the embryos immediately before transfer, as the state of the embryo at earlier times provides no additional information.

Taken together, our results show that the development of oocytes and embryos emerges as a simple process, despite the underlying molecular complexities of the biology and despite the plethora of disease aetiologies and treatment protocols presenting in a clinic. More broadly, this work validates the use of phenomenological models of oogenesis and embryogenesis by demonstrating that simple models can be constructed without sacrificing quantitative accuracy. Although we infer the models using data drawn from controlled ovarian stimulation and not from natural menstrual cycles, the models provide insight into the principles that govern oogenesis and embryogenesis, and may be useful in guiding clinical IVF treatments.

2. Results and discussion

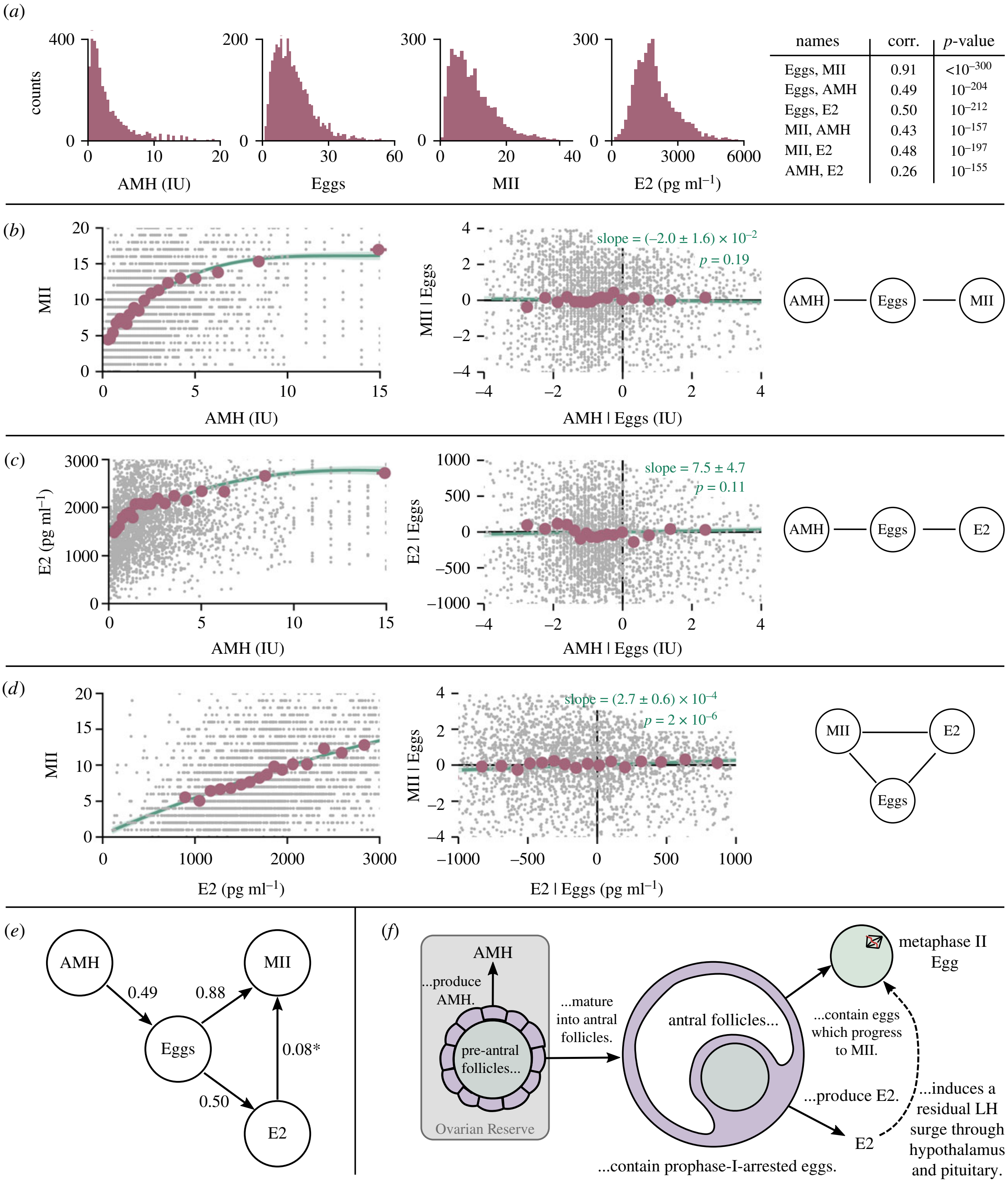
2.1. A model of oocyte development

The formation of a healthy embryo begins with the successful progression of an oocyte from prophase-I arrest to metaphase-II. Which clinical factors affect the number of metaphase-II oocytes retrieved during an IVF ovarian stimulation?

The clinical data contain 69 variables that describe either the patient or the ovarian stimulation. We start by examining four variables that are strongly correlated with the number of metaphase-II oocytes: (1) the number of total eggs retrieved during an ovarian stimulation cycle (Eggs), (2) the number of eggs in metaphase-II arrest (MII), (3) the patient's maximum serum oestradiol concentration during the cycle (E2) and (4) the patient's serum AMH before the cycle. Oestradiol is a hormone produced by the ovaries during natural and stimulated ovulatory cycles [30]; AMH is considered a measure of the patient's ovarian reserve [26]. Each of these variables varies widely across the 4910 cycles for which all four variables are recorded, with coefficients of variation of 0.5–1.2 (figure 1*a*, left). All four variables are strongly correlated with one another, with Pearson correlation coefficients between 0.26 and 0.91 and *p*-values between 10^{-300} and 10^{-155} (figure 1*a*, right).

To understand which factors quantitatively affect oocyte maturation, we first search for conditional independencies among the variables AMH, MII and Eggs. A conditional independency between two variables implies that one variable can be completely described without direct knowledge of the other, suggesting the existence of a simple phenomenological model. To search for conditional independencies, we nonlinearly regress both MII on Eggs and AMH on Eggs, by finding the best-fit polynomial that maximizes the Bayesian posterior evidence. This method allows for capturing complex dependencies without overfitting the data [31,32] (see electronic supplementary material, §1); we also split the data into separate train and test sets as a further check against overfitting. We then take the residuals from the two regressions and evaluate their correlation. We denote this procedure as Corr(AMH, MII | Eggs). We find that, although there is a strong correlation between AMH and MII (figure 1*b*, left), that correlation disappears after conditioning on Eggs: Corr(AMH, MII, | Eggs) = −0.02 ($p = 0.19$; figure 1*b*, centre). This correlation is both consistent with zero and smaller than an effect size threshold of 0.05, suggesting that AMH and MII are conditionally independent given Eggs.

We encode this conditional independency using a class of graphical models known as Bayesian networks. These have found usage in causal inference [33–36]; here, we use them to construct phenomenological models that correspond to



names	corr.	p-value
Eggs, MII	0.91	$<10^{-300}$
Eggs, AMH	0.49	10^{-204}
Eggs, E2	0.50	10^{-212}
MII, AMH	0.43	10^{-157}
MII, E2	0.48	10^{-197}
AMH, E2	0.26	10^{-155}

Figure 1. (*a*) Distributions and correlations of the four variables AMH (measured in International Units, IU), Eggs, MII and E2 (measured in pg ml^{-1}). (*b*) Left: the number of metaphase-II oocytes retrieved (MII) is strongly correlated with the patient's serum AMH. Grey dots: raw data, red circles: raw data binned into 20 separate bins with equal counts, green line and shaded region: nonlinear regression and errors. Centre: that correlation disappears after regressing against the total number of retrieved oocytes (Eggs). Grey dots: residuals after regressing against Eggs, red circles: residuals binned into 20 separate bins with equal counts, green line and shaded region: linear fit to the residuals, with slope and standard error shown at top of plot. The axes range of the plots is cropped to show details of the trends. Right: this conditional independency suggests a graph of the form AMH–Eggs–MII. (*c*) Left: the patient's serum oestradiol concentration (E2) is strongly correlated with AMH. Centre: that correlation disappears after regressing against Eggs. Right: this conditional independency suggests a graph of the form AMH–Eggs–MII. (*d*) Left: MII is strongly correlated with E2. Centre: while regressing against Eggs greatly weakens that correlation, E2 and MII remain correlated after conditioning on Eggs. This suggests a fully connected graph is needed to describe these three variables (right). (*e*) A graphical model that is consistent with the data. Edge labels show the conditional correlation coefficients after conditioning on all other incoming edges; the data are consistent with the arrow marked with a * oriented in either direction. (*f*) The graphical model expected from prior knowledge of ovarian stimulation.

mechanistic descriptions of biology. Briefly, for a given factorization of a probability distribution, these graphs contain a directed edge from one variable to another if the probability of the second explicitly depends on the first. Two variables are conditionally independent if all paths from one variable to the other are 'blocked', by head-to-tail or tail-to-tail nodes meeting at a variable that is conditioned on, or by head-to-head nodes meeting at an edge that is not conditioned on [34,37]. Both the observed correlation between MII and AMH and their conditional independency given Eggs can be captured by any of the graphs AMH → Eggs → MII, AMH ← Eggs → MII, or AMH ← Eggs ← MII; we denote this ambiguity by AMH–Eggs–MII, with an as-yet undetermined orientation of the arrows (figure 1*b*, right). If the data are described by one of these graphs, then Eggs and AMH should remain correlated given MII, which is indeed the case: Corr(Eggs, AMH | MII) = 0.24 ($p = 10^{-47}$; electronic supplementary material, figure 2*a*). Likewise, MII and Eggs should remain correlated given AMH, which is the case: Corr(MII, Eggs | AMH) = 0.87 ($p < 10^{-300}$; electronic supplementary material, figure S2*b*).

Second, we examine the variables AMH, E2 and Eggs. AMH and E2 are strongly correlated (figure 1*c*, left), but regressing AMH and E2 on Eggs shows that Corr(AMH, E2 | Eggs) = 0.03, consistent with no conditional correlation ($p = 0.11$; figure 1*c*, centre). This suggests a graph of the form AMH–Eggs–E2 (figure 1*c*, right). Third, we examine Eggs, MII and E2. While regressing on Eggs greatly weakens the correlation between MII and E2 (compare figure 1*d* left and centre), the measured conditional correlation is still positive: Corr(E2, MII | Eggs) = 0.08, $p \approx 10^{-6}$. Thus, an edge must connect each of Eggs, E2 and MII (figure 1*d*, right). Finally, we examine the variables AMH, E2 and MII and find no additional conditional independencies (electronic supplementary material, figure S2).

Of the 543 graphical models that describe four variables, only eight models capture exactly the two conditional independencies described above. The data alone cannot distinguish between these graphs. However, the patient's AMH is measured before the ovarian stimulation starts, whereas the other three variables are measured during the treatment. Thus, any graph with an edge pointing into AMH cannot correspond to a mechanistic description of the biology. Ruling out these graphs leaves only two graphs consistent with both the data and a mechanistic interpretation (figure 1*e*).

Physiologically, this quantitative, phenomenological model recapitulates our qualitative understanding of ovarian stimulation (figure 1*f*): (1) pre-antral follicles, containing immature oocytes and associated somatic cells, produce the hormone AMH [23,25,38]. Since pre-antral follicles can grow into large antral follicles with prophase-I-arrested eggs, AMH is a measure of the potential number of oocytes that could develop. This is captured by the inferred arrow in figure 1*e* from AMH to Eggs, which indicates that the patient's AMH determines how many eggs she will produce. (2) Antral follicles produce oestradiol, captured by the inferred arrow from Eggs to E2. (3) Some, but not all, eggs progress from prophase-I arrest to metaphase-II arrest. This is captured by the inferred arrow from Eggs to MII. (4) During natural ovulation, the oestradiol produced by antral follicles signals the pituitary and hypothalamus to release hormones which modulate oocyte maturation. During an ovarian stimulation, clinicians attempt to temporarily disable this feedback between the hypothalamus and the pituitary [30], suggesting that oestradiol should not impact oocyte maturation during ovarian stimulation. However, the inferred arrow from E2 to MII suggests that a weak feedback between the hypothalamus, the pituitary and oestradiol is still present during an ovarian stimulation cycle.

This inferred phenomenological model (figure 1*e*) provides a quantitative representation of oocyte development that allows direct and indirect effects to be disentangled. For instance, a patient starting treatment with a higher AMH is likely to produce more oocytes, via the direct arrow AMH → Eggs. In addition, that patient is likely to have a higher oestradiol level during the cycle, as the additional eggs she is likely to produce will on average produce more oestradiol, via the path AMH → Eggs → E2. However, the graph states that this effect is indirect: the patient's E2 increases only through the associated increase in the number of eggs for high-AMH patients. This is borne out by the data. Likewise, a patient with a larger number of MII oocytes retrieved is likely to have a higher AMH, since following the arrows backwards shows that higher MII implies that Eggs is higher, and higher Eggs implies a higher AMH. However, once again, this is an indirect effect; the patient's AMH is more likely to be high only because of the associated increase in Eggs when many MII oocytes are retrieved.

2.2. Oocyte maturation is robust to other factors

A woman's age and obesity are known to affect her fertility and IVF prognosis. Does age or obesity directly affect oocyte maturation after accounting for E2 and Eggs?

The data show that age has no direct effect on an oocyte's ability to reach MII. The conditional correlation between MII and age is consistent with zero: Corr(Age, MII | Eggs, E2) = 0.02 ($p = 0.35$, figure 2*a*). In addition, the data constrain the magnitude of any effect to be tiny. Fitting a line to the residuals constrains the slope to be $(1.1 \pm 1.2) \times 10^{-2}$ MII oocytes/year (point estimate ± standard error). To place this in perspective, consider a treatment cycle for two women, aged 33 and 40 years old (the 25th and 75th percentile in the data). If the treatment results in the same number of total oocytes and the same max oestradiol for both women, then on average the number of retrieved MII oocytes should differ by no more than 0.2. Since the median MII oocytes retrieved per cycle is eight, the data constrain the direct effect of age on MII to be 2% or less.

Likewise, the data show that the woman's BMI has no direct effect on an oocyte's ability to reach MII. The conditional correlation between MII and BMI is consistent with zero: Corr(BMI, MII | Eggs, E2) = −0.005 ($p = 0.78$; figure 2*b*). The data constrain the direct effect of BMI on MII to be 1% or less.

Ovarian stimulation drugs are necessary for multiple oocytes to reach metaphase-II in a single cycle. However, the data show that the dose of these drugs has no direct effect on an oocyte's ability to reach MII. The conditional correlation between MII and the dose of FSH or HMG is consistent with zero: Corr(FSH, MII | Eggs, E2) = 0.01 ($p = 0.49$; figure 2*c*), Corr(HMG, MII | Eggs, E2) = −0.03 ($p = 0.08$; figure 2*d*). The data constrain the direct effect of these stimulation drugs to be 1% or less for FSH, and 5% or less for HMG. These observations show that oocytes develop to MII independently of a patient's age, her BMI, or details of the ovarian stimulation procedure.

Perhaps only a fixed number of oocytes per cycle are capable of maturing to metaphase-II, and therefore the probability of an oocyte being metaphase-II depends on the total number of eggs retrieved. However, while Eggs is strongly

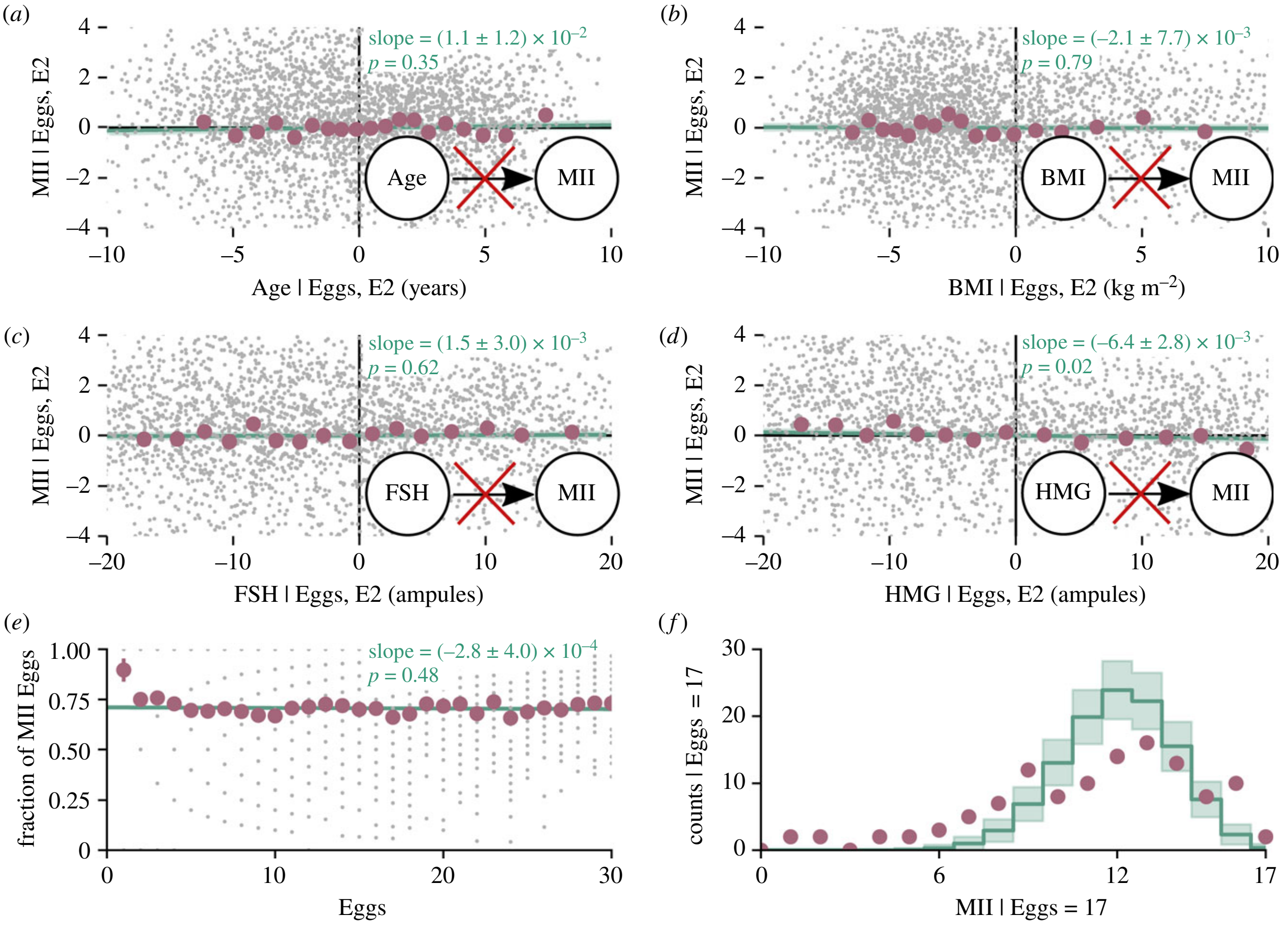


Figure 2. MII versus the patient's age (*a*), BMI (*b*), and the doses of stimulation drugs FSH (*c*) and HMG (*d*), after regressing against Eggs and E2. Grey dots show the residuals from the regressions, red circles and error bars show the mean and standard error of the data binned into 20 bins with equal number of points, and green lines, shaded regions and labelled slopes show the mean and standard error of the best linear fit to the residuals. (*e*) The fraction of MII oocytes (MII/Eggs) versus Eggs. (*f*) Histogram of observed MII (red circles) versus that expected from independently triggering follicles (green line and shaded region show the expected counts and their standard deviation), for the 116 cycles that have 17 eggs retrieved, which is where the discrepancy between the two histograms is the largest as measured by a χ^2-test ($p < 10^{-300}$, primarily due to the cycles with MII of 1–5). On the scale of the plot, the expected histograms from a binomial distribution where the probability for an oocyte being metaphase-II is constant is indistinguishable from one where the probability varies with E2.

predictive of MII, it provides no predictive power for the fraction of eggs in MII arrest (MII/Eggs): Corr(MII/Eggs, Eggs) = −0.01 ($p = 0.48$). Linearly regressing MII/Eggs on Eggs gives a slope tightly constrained near zero (figure 2*e*).

Combined, these observations suggest the following simple picture for oocyte maturation: each follicle independently triggers its oocyte to leave prophase-I, with a probability that depends only on E2. The oocyte then progresses to metaphase-II, with both processes independent of interactions with other follicles or the aggressiveness of the ovarian stimulation.

To check whether this simple picture completely describes oocyte maturation, we examine the distribution of MII for fixed Eggs. If each follicle independently triggers its egg to progress to metaphase-II with some probability p, then MII for each cycle will be binomially distributed, denoted as B(MII; Eggs, p). If that probability depends only on E2, then the measured distribution of MII across cycles with a given Eggs should be the average of many binomial distributions, each with a probability that depends on E2: $\langle B(\mathrm{MII};\ \mathrm{Eggs},\ p(\mathrm{E2}))\rangle_{\mathrm{E2}}$. Instead, the empirical distribution is much broader than the expected one (figure 2*f*). This discrepancy suggests that additional factors affect oocyte maturation, such as biochemical processes within the oocyte that are shared by multiple eggs from the same patient, interactions between follicles beyond serum oestradiol, or simply other clinical factors that we have not accounted for.

The data provide some insight into human meiosis, especially given prior knowledge of human ovulation. As a woman ages, the oocytes she ovulates become much more likely to be aneuploid, rising from an aneuploidy rate of roughly 25% at age 30–80% at age 42 [27,39,40]. Chromosomal signatures show that aneuploidy in human oocytes can arise both during the oocyte's progression from prophase-I to metaphase-II and immediately after fertilization [11,41–43]. In mitotic cells, mis-segregation of chromosomes is reduced by the spindle-assembly checkpoint [44], which can arrest mitosis until chromosomes are properly aligned. If there were a strong spindle assembly checkpoint in meiosis I, then the typically aneuploid oocytes from older women would reach metaphase-II at a reduced incidence than those from younger women. Instead, oocytes from older and younger women reach metaphase-II arrest at the same incidence. This is consistent with experimental work that shows that human oocytes have a weak meiotic spindle assembly checkpoint [45–47].

2.3. Follicular recruitment is robust to patient factors

While the patient's age, BMI and dose of ovarian stimulation drugs do not directly affect an oocyte's progression to metaphase-II, they could affect the rest of follicular recruitment. To investigate this, we examine the joint distribution of all of Age, BMI, AMH, FSH, HMG, Eggs, E2 and MII, constructing a directed acyclic graph that is consistent with the data. The number of possible graphs grows rapidly with the number of variables: there are 543 possible graphs for four variables, but 783, 702, 329, 343 possible graphs with eight variables. To deal with this inordinately large number of possible models, we use prior knowledge to split the variables into three groups: prognostic variables measured before the treatment starts (Age, BMI and AMH), treatment variables (FSH and HMG), and response variables measured after the drugs have been applied (Eggs, MII, E2). We then search for graphs that are consistent with a mechanistic interpretation, by excluding graphs with edges directed from treatment to prognostic variables, from response to prognostic variables, or from response to treatment variables.

Among the prognostic variables, the data show that BMI does not directly affect AMH: Corr(AMH, BMI | Age) = −0.04 ($p = 0.01$; electronic supplementary material, figure S4*a*). Physiologically, this implies that the likelihood that a primordial follicle develops into a pre-antral follicles is independent of obesity. This is the only conditional independency among the prognostic variables.

The data show that clinicians customize the doses of FSH and HMG based on the patient, as expected. The FSH dose depends on all three prognostic variables, and the HMG dose depends on Age, AMH and FSH, but not BMI (electronic supplementary material, figure S4*f*). The lack of conditional independencies between the treatment and prognostic variables demonstrates that the conditional independencies we do see elsewhere are real and not an artefact of our analysis.

Among the response variables, the data show that follicular recruitment is simple, although follicular hormone production is not. The data are consistent with Age and BMI having no direct effect on Eggs: Corr(Age, Eggs | AMH), HMG = −0.04, $p = 0.01$; Corr(BMI, Eggs | AMH, HMG) = −0.01, $p = 0.54$ (electronic supplementary material, figure S4*c*,*d*). Physiologically, this suggests that the ability of a pre-antral follicle to be recruited does not worsen with age or obesity. Likewise, Eggs is conditionally independent of FSH: Corr(FSH, Eggs | AMH, HMG) = −0.04 ($p = 0.02$; electronic supplementary material, figure S4*e*). This suggests that, at this particular clinic, clinicians prescribe sufficient FSH to recruit all the follicles in the cohort activated from the primordial pool in that menstrual cycle. Interestingly, we observe a weak negative correlation between Eggs and the dose of HMG: Corr(HMG, Eggs | AMH) = −0.11 ($p = 10^{-11}$). Taken at face value, this seems to imply that HMG is typically supplied at more than the optimal dose at this clinic. While there is some evidence that excessive HMG can cause follicles to degrade [48,49], another possibility is that the negative correlation is due to clinicians prescribing more HMG to patients whom they know *a priori* to be poor responders even after accounting for their age, BMI and AMH—for example, patients who have had a poor response in previous treatments. However, the correlation remains negative even when excluding patients on repeat stimulation cycles: Corr(HMG, Eggs | AMH, first cycle) = −0.09, $p \approx 10^{-5}$. Combined, the data paint a simple picture for follicle recruitment: all available follicles are typically recruited, independently of effects from age or obesity but weakly affected by HMG. Finally, in contrast to the simplicity of Eggs and MII, all of Age, BMI, FSH, HMG and Eggs affect E2. These observed conditional independencies are captured by the graphical model in figure 3*a*.

The model in figure 3*a* predicts 99 conditional independencies among the eight variables shown. If the model completely describes the data, then the 99 corresponding conditional correlations must be consistent with zero, i.e. none of their *p*-values should be statistically significant. As a stringent check of the model, we measure the conditional correlations and associated *p*-values for each of these 99 independencies in both the train and test sets. We then check that the measured conditional correlations are consistent with zero by comparing their associated *p*-values to those calculated from datasets simulated according to the model in figure 3*a*. We find that the measured *p*-values are consistent with the simulated ones, although the lowest measured *p*-value is lower than the typical simulated one (figure 3*b*). By contrast, datasets generated according to more complicated models display a different distribution of these *p*-values (figure 3*c*). Moreover, anything missing from the model in figure 3*a* must correspond to a small effect with small explanatory power. Fitting the training data with a fully-connected model explains 0.6% or less of each variable's variance in the test data, with the fully-connected model actually performing worse on the test set for most of the fits than the model in figure 3*a* does (electronic supplementary material, §4). Combined, these observations show that the graphical model accurately describes human ovarian physiology and oogenesis.

Viewed holistically, the probabilistic graphical model in figure 3 has a simple mechanistic interpretation. The patient's age and BMI only act to determine the hormone levels. Hormones other than oestradiol determine how many antral follicles develop. Follicles then produce oestrogen and trigger eggs to progress to metaphase-II, with a slight feedback between these. Because of this simplicity, changes due to age or obesity manifest themselves in simple ways, after ignoring the physiologically irrelevant question of how clinicians choose drug doses for ovarian stimulation. In particular, the only direct effect of obesity on the oocyte maturation process is to decrease E2. This is consistent with work showing that obesity affects fertility by interfering with the hormonal regulation of ovulation [29]. The primary effect of age on the oocyte maturation process is to decrease the ovarian reserve, as measured by the patient's AMH. Physiologically, this is consistent with the well-known decrease in a woman's pool of primordial follicles as she ages [30].

While this model is a quantitative description of oocyte maturation, it is not a microscopic model. There are many intermediate factors that are not included in the model but are known to affect follicular recruitment and oocyte maturation, such as GnRH from the hypothalamus, FSH and LH from the pituitary and inhibin, androstenedione and cAMP from the follicles. However, the model's accuracy suggests that these additional factors are intermediate and can be coarse-grained out to give a macroscopic, phenomenological description of oocyte maturation. Moreover, the model's accuracy also shows that these additional factors do not cause additional dependencies between variables: the model places strong constraints on age-related sensitivity of oocytes to the maturation signal, for example.

AMH and E2 are produced by ovarian follicles at different stages in follicular growth. The data provide insight into both

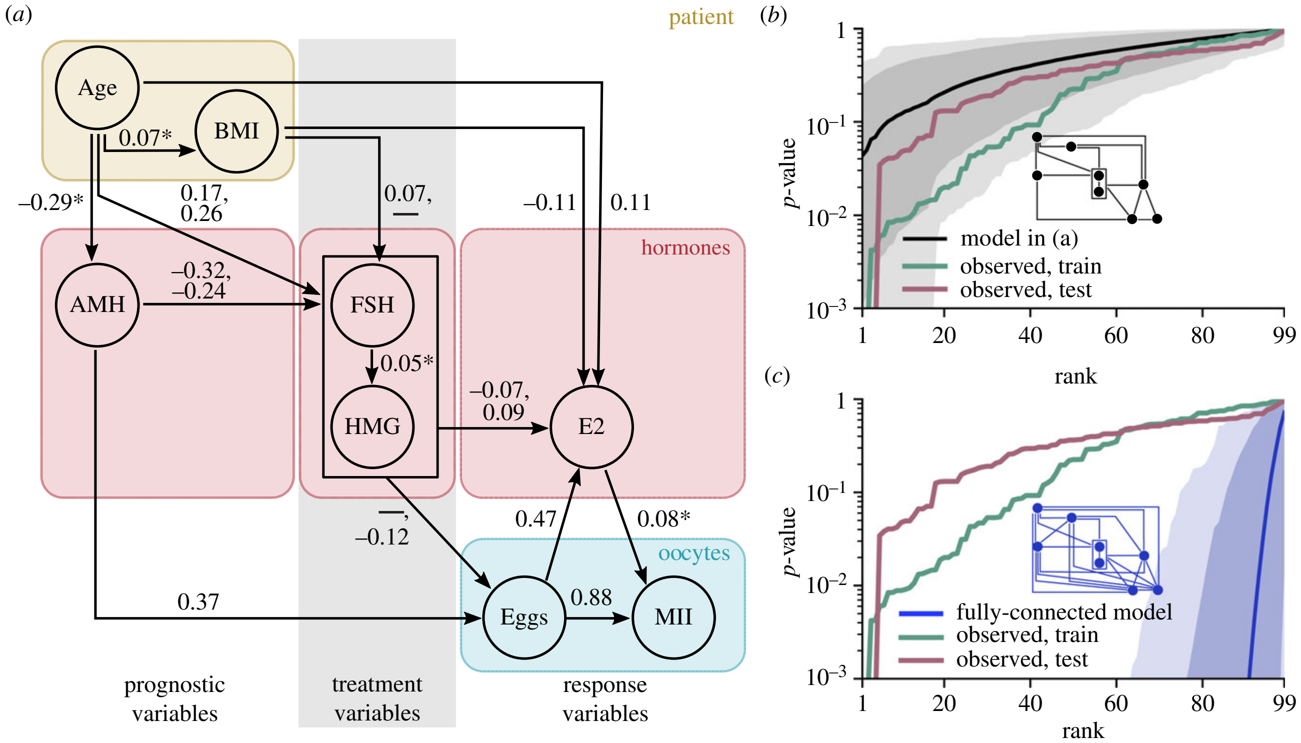


Figure 3. (*a*) A graphical model of oogenesis that is consistent with a mechanistic interpretation of the data. Labels show conditional correlation coefficients. For edges with two labels, the upper corresponds to FSH, the lower to HMG; dashes signify no dependence. The data are consistent with marked arrows (*) oriented in either direction. (*b*) Rank plots of *p*-values for the 99 conditional correlations corresponding to the conditional independencies predicted by the model. The green line shows that from the training data, the red line that from the test data. The black line and shaded regions show the median, 95% and 99.9% centred percentile of rank plots from 3000 datasets simulated according to the proposed model. (*c*) The same as (*b*), but showing the distribution of rank plots for the 99 conditional correlations from fully connected, linear Gaussian models.

these processes. For a given number of retrieved oocytes, the mean of AMH is linear in the number of retrieved oocytes, with an intercept near zero (figure 4*a*, also electronic supplementary material, figure S5). For Eggs not too large, the variance of AMH is also linear in Eggs (figure 4*b*). The deviation from linearity at large Eggs is largely due to patients with polycystic ovary syndrome, who tend to have high AMH; excluding patients with a diagnosis of ovulatory dysfunction (i.e. polycystic ovary syndrome) brings most of the Var(AMH) measurements onto the linear fit (data not shown). Since means and variances add when summing independent random variables, the linearity of both the mean and the variance of AMH in the number of follicles suggests that each follicle produces AMH independently from interactions with other follicles. (The small but non-zero intercept could arise if some pre-antral follicles do not mature sufficiently to be retrieved during the IVF retrieval procedure.) By contrast, the mean E2 is not linear in Eggs, systematically deviating from linearity and having an intercept that is far from zero (figure 4*c*; variance in panel *d*). These variations from linearity show that the E2 is not produced independently by each follicle, perhaps due to additional production from outside follicles (such as by adipose tissue or the adrenal glands), due to inter-follicle feedback, or simply due to other factors that affect E2 production, such as those shown in figure 3*a*.

2.4. Pre-implantation development

Once the oocyte is fertilized, the resulting embryo starts to divide. By the third day after fertilization, a human embryo typically has eight cells. As its cells continue to divide, on the fifth day the embryo differentiates into a blastocyst, composed of two distinct cell lineages: the trophectoderm and the inner-cell mass. In natural development, the blastocyst then attaches to the woman's endometrial epithelium and implants in her uterus [10,12,30,50]. In the IVF clinic, human embryos are typically cultured for 3 or 5 days after fertilization, at which point embryologists attempt to select the highest quality embryo(s), which is then transferred into the patient's uterus.

We first examine the overall trajectory of development, as described by three variables: the number of cells in the embryo on Day 3 after fertilization (Day 3 Cells), its developmental stage on Day 5 (Day 5 Stage), and whether it resulted in a fetal heartbeat after transfer (FH; see electronic supplementary material, SI for details). Day 5 stage is scored from 1 to 9: (1) degenerate or arrested; (2) morula with incomplete compaction, (3) morula with complete compaction, (4) early blastocyst, (5) expanding blastocyst, (6) full blastocyst, (7) expanded blastocyst, (8) hatching blastocyst and (9) hatched blastocyst.

However, not all embryos are cultured to Day 5, and thus not all embryos have data on both Day 3 and Day 5: of the 55 350 embryos recorded on Day 3, only 41 932 are also recorded on Day 5. Since the decision to culture embryos to Day 5 is made based on patient prognosis and embryo quality, embryos that are assessed on Day 5 systematically differ from those that are not. To avoid biases due to this missing data, we treat missingness as an additional variable and model both the variable and its missingness [51]. The clinical data are in the 'missing at random' regime, where a

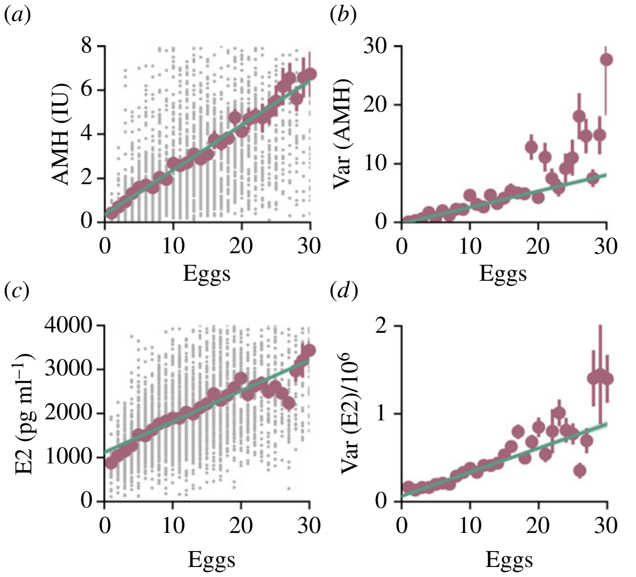


Figure 4. (*a*) AMH versus Eggs. Grey dots show raw data, red circles show the mean and standard error of AMH binned at each value of Eggs and green line shows the best linear fit to the data, with slope of 0.20 ± 0.01 and intercept 0.36 ± 0.11. A linear fit is the best-fit polynomial to the data, as determined by the model evidence. (*b*) Variance and error estimate of AMH versus Eggs, trimmed to the central 95% for each value of Eggs (red dots, errors calculated using the variance of the *k*-statistic). The green line shows the best linear fit to the variance, with a slope of 0.28 ± 0.01 and intercept −0.22 ± 0.03. (*c*) E2 versus Eggs. The green line shows the best linear fit to the data. A quadratic model (not shown) provides the best fit to E2 versus Eggs. (*d*) Variance and error estimate of E2 versus Eggs, trimmed to the central 95% for each value of Eggs.

variable's missingness depends on other variables in the dataset, but not on the missing variable itself. In the language of probabilistic graphical models, there are edges from some of the normal variables to the missingness variables, but no edge from a variable to its own missingness. In this regime, valid inferences require conditioning on missingness and the variables on which the missingness depends. Multiple transfers cause an additional problem for the measurement of fetal heartbeat—if two embryos are transferred simultaneously and one fetal heartbeat is observed, it is not obvious which embryo formed the fetus. We solve this problem via generative modelling. Briefly, we construct a parameterized model that predicts the probability of one embryo implanting from properties of the embryo and the woman. We then fit the model to the data by finding the maximum *a posteriori* parameters, using a Poisson-binomial likelihood for multiple transfers. To check whether a variable is conditionally independent of fetal heartbeat, we fit two models, one with the additional variable and one without, and perform Bayesian model selection to see if the additional variable is necessary (electronic supplementary material, SI §1). We include all available cycles with four or fewer embryos transferred (95% of cycles).

With this approach, we construct a model of human pre-implantation development. The embryo's number of cells on Day 3 is strongly correlated with its stage on Day 5: Corr (Day 3 Cells, Day 5 Stage |both measured) = 0.44 ($p < 10^{-300}$, figure 5*a*). Both the embryo's Day 3 Cells and its Day 5 Stage are individually predictive of fetal heartbeat (figure 5*b*,*c*), as appreciated in the literature [52–54]. However, when considered jointly, only Day 5 stage is predictive of fetal heartbeat; Day 3 Cells provides no additional information whether the embryo will develop (figure 5*d*). This conditional independence implies a model of the form Day 3 Cells → Day 5 Stage → FH; this is the only graph with two or fewer edges that is consistent with the data and the fact that Day 3 happens before Day 5.

How do the woman's age, her BMI and the aggressiveness of the ovarian stimulation additionally affect her embryos' development? The woman's age affects all stages of her embryos' development, being weakly correlated with the embryo's number of cells on Day 3 and its stage on Day 5: Corr(Day 3 Cells, Age) = −0.07 ($p = 10^{-47}$; electronic supplementary material, figure S6*a*), Corr(Day 5 Stage, Age | Day 3 Cells, measured) = −0.11 ($p = 10^{-78}$; electronic supplementary material, figure S6*b*), and having a strong effect on the probability that an embryo forms a fetal heartbeat (electronic supplementary material, figure S7*c*). Surprisingly, the woman's BMI affects none of Day 3 Cells (electronic supplementary material, figure S6*c*), Day 5 Stage (electronic supplementary material, figure S6*d*) or FH (electronic supplementary material, figure S7*e*). Likewise, the conditional correlation of MII with all of D3 Cells, D5 Stage and FH is either consistent with zero or less than 0.05 in magnitude (electronic supplementary material, figures S6*e*,*f* and 7*f*). Combined, these observations yield the simple graphical model for development in figure 5*e*. The woman's age determines the embryo's cell number on Day 3; her age and the embryo's cell number on Day 3 determine its stage on Day 5; and her age and the embryo's stage on Day 5 determine whether it will continue to develop. Neither the patient's BMI nor the number of retrieved MII eggs significantly affect the embryo's development. By contrast, the data's missingness shows a much more complex distribution, with almost all variables affecting the clinical decisions regarding embryo transfer and culture duration. Once again, the data paint a picture of simple biology but complex clinical decisions.

That Day 3 Cells has no direct effect on the fetal heartbeat probability is particularly striking. The time-varying morphology (morphokinetics) of human embryos is known to be predictive of developmental success [30,53]. In principle, morphokinetics up to Day 3 could provide a different set of information than morphokinetics between Days 3 and 5. For example, since the embryo's genome activates on Day 3 [55], one might reasonably propose that the embryo's progress before Day 3 provides information about the ooplasm, that the embryo's progress from Day 3 to Day 5 provides information about aneuploidy, and that both of these factors independently determine the embryo's prognosis. Instead, the data show that the embryo's stage at Day 5 contains all the information about its viability that its cell number at Day 3 contains.

One physiological interpretation consistent with the data is that there are two distinct sets of mechanisms that influence an embryo's development (figure 5*f*). One set of mechanisms influences pre-implantation development only through its overall rate, including the number of cells on Day 3 and the stage on Day 5. This set of mechanisms depends weakly on age. Another set of mechanisms influences an embryo's post-implantation developmental potential and depends strongly on age. Perhaps surprisingly, meiotic aneuploidy of the oocyte cannot strongly affect pre-implantation development, since meiotic aneuploidy is strongly associated with the woman's age. Instead, the data suggest that meiotic

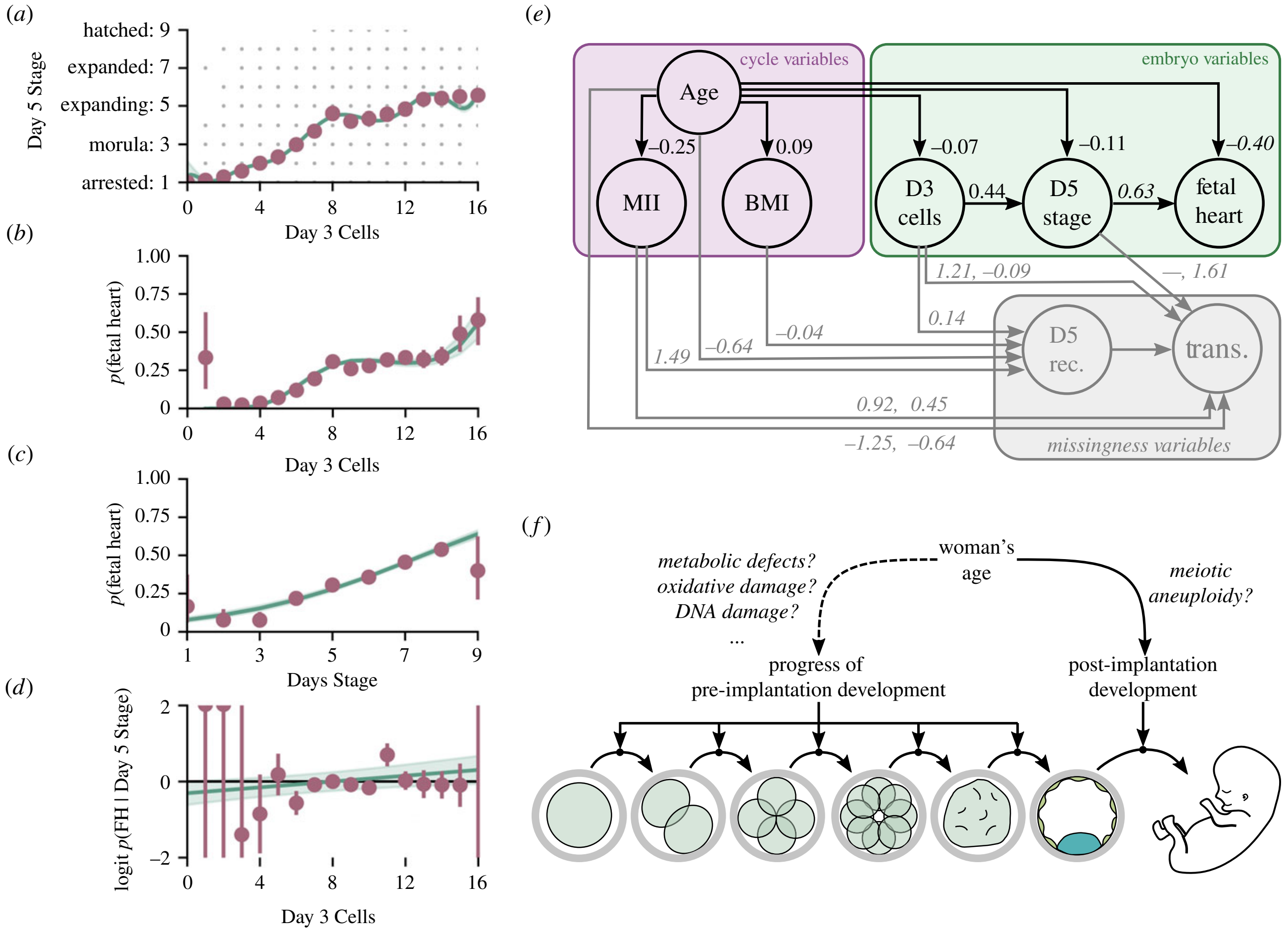


Figure 5. (*a*) Day 5 Stage versus Day 3 Cells. Grey dots show the raw data, red circles show the mean Day 5 Stage for each separate value of Day 3 Cells, the green line and shaded region show the nonlinear model with the highest evidence and its uncertainty. The Day 5 Stages are: (1) degenerate or arrested, (2) morula with incomplete compaction, (3) morula with complete compaction, (4) early blastocyst, (5) expanding blastocyst, (6) full blastocyst, (7) expanded blastocyst, (8) hatching blastocyst, and (9) hatched blastocyst (see electronic supplementary material, §2 for details). (*b*) Estimated probability of an embryo resulting in a fetal heartbeat (FH) as a function of Day 3 Cells alone, for embryos recorded on Day 3 and transferred. The red circles and error bars show the probability estimated by a model that fits an independent probability of implantation for each number of cells; the green line and shaded region shows the nonlinear model with the highest model evidence and its uncertainty. (*c*) The estimated probability of FH as a function of Day 5 Stage alone, for embryos recorded on Day 5 and transferred. (*d*) The logit of the estimated probability of FH as a function of Day 3 Cells, after regressing against Day 5 Stage. Red circles and error bars show the additional log probability estimated from a model that fits an independent logit for each value of Day 3 Cells; green line shows the best linear model and uncertainty for the logit. The data are consistent with Day 3 Cells having no additional predictive power on FH once Day 5 Stage is known. (*e*) The graph with the minimal number of edges that is consistent with the data and a mechanistic interpretation. Black nodes and arrows show the measured data; grey nodes and arrows show the data's missingness. Only the arrow Age → BMI can be re-oriented without breaking consistency with the data or a mechanistic interpretation. Edge labels for continuous variables are conditional correlation coefficients (roman typeface). Edge labels for discrete variables are coefficients from logistic regression (italic typeface), after treating the effects of other variables with edges into the discrete variable and after normalizing the input variable by its mean and standard deviation. The missingness variables D5 Rec. and Trans. are 1 if the embryo is recorded on Day 5 or transferred, respectively, and 0 otherwise. The distribution of Trans. changes depending on whether Day 5 Stage was recorded; the two labels on edges into Trans correspond to Day 5 Stage missing or recorded. (*f*) The data are consistent with a picture where processes which control pre-implantation development are largely different from those which control post-implantation development.

aneuploidy primarily affects post-implantation, rather than pre-implantation, development. Other works have provided mixed evidence regarding the extent to which aneuploidy affects pre-implantation development [56–59].

Is development genuinely this simple, or is this apparent simplicity an artefact of the variables we chose to describe the embryo? In addition to the number of cells, the dataset also describes the embryo on Day 3 with the presence of cytoplasmic fragments, the presence of multiple nuclei in individual cells, size asymmetries among cells within the embryo, the presence of large vacuoles, and the granularity of the cell cytoplasm. On Day 5, the dataset also includes grades of the inner-cell mass and the trophectoderm, for embryos that have formed blastocysts. Of the Day 3 variables, embryo fragmentation and cell symmetry are predictive of fetal heartbeat, along with Age and Day 3 Cells. Likewise, of the Day 5 variables, the trophectoderm grade is predictive of fetal heartbeat, along with Age and Day 5 Stage, in agreement with recent studies [60,61]. (The data are consistent with the other Day 3 variables providing no additional predictive power once Age, D3 Cells, symmetry, and fragmentation are known; and the data are consistent with the inner-cell mass grade providing no additional predictive power once Age, D5 stage and the trophectoderm grade are known.) Nevertheless, the Day 3 variables provide no additional predictive power for fetal heartbeat once the Day 5 variables are known (electronic supplementary material, §4). These observations suggest a memoryless model of pre-implantation development: provided the embryo makes it to the blastocyst stage, what happened before is irrelevant for its viability. Thus, despite the molecular complexities of early development and the complicated trajectory of human development before 12 weeks, a simple, phenomenological view of embryonic viability may be possible without sacrificing quantitative accuracy.

3. Conclusion

Here, we have used clinical IVF data and minimal prior knowledge to infer quantitative, phenomenological models of human oogenesis and embryogenesis. Not only does constructing these models with a data-driven approach give confidence in their validity, but the models recapitulate known aspects of oogenesis and embryogenesis. Surprisingly, the models that best describe the data are sparse, with only one or two factors affecting most physiological processes. This suggests that oogenesis and embryogenesis are modular processes. Our analysis leads to three additional, surprising conclusions which support this overall picture of modularity.

(i) AMH production by one pre-antral follicle is independent of the amount produced by others. This is in stark contrast to other follicularly produced hormones. Hormones such as oestradiol and inhibin participate in feedback loops which regulate the formation of the dominant follicle in a natural cycle. As a result, these hormones are produced in a highly regulated manner, and not independently by each follicle [30,62]. Moreover, mathematical modelling suggests that, for these feedback loops to function, the hormone production and response needs to be highly nonlinear [63–65]. Thus, the amount of these hormones produced by one follicle depends strongly on the amount produced by other follicles. By contrast, AMH appears to be produced without regulatory feedback. This is particularly interesting because AMH regulates the follicle number, by regulating the growth of primordial follicles into primary (pre-antral) follicles. Thus, while feedback loops are needed to accurately control the number of mature follicles recruited during natural ovulation, feedback loops appear to be unnecessary to sufficiently control the number of primary follicles recruited.

(ii) Neither age, obesity, the ovarian stimulation, nor even the number of recruited oocytes affects whether an individual oocyte progresses to metaphase-II once triggered to resume meiosis. These observations have several implications for the biology of the oocyte. Definitive evidence shows that age is strongly correlated with aneuploidy in the oocyte [27,39]. Since our analysis shows that the ability of an oocyte to progress to metaphase-II is independent of age, we conclude that this ability is the same for both euploid and aneuploid oocytes. Thus, the spindle assembly checkpoint in human oocytes must be weak, in agreement with recent experimental work [45–47]. A similar argument can be made regarding the effect of metabolism on meiosis. Evidence suggests that oocyte mitochondrial metabolism worsens with increasing obesity [29,66]. Since oocytes progress to metaphase-II independently of obesity, metabolic defects must not typically be enough to stop an oocyte from progressing from prophase-I to metaphase-II.

(iii) Early embryonic development is memoryless, in that embryos with the same status on Day 5 develop the same, regardless of their status on Day 3. This memorylessness is reminiscent of the robustness of early mammalian embryos to damage to individual cells [67–69]; however, memorylessness is more than robustness. Robustness signifies that an embryo can recover from a setback. Memorylessness signifies that, once recovered, neither the setback nor what caused it has any impact on the rest of development. The memorylessness implies a modularity and robustness in development, similar to previous proposals for modularity in cellular biology [70].

The models we present also have implications for clinical IVF. For embryos transferred on Day 5, embryo selection can be based solely on how developed they are on Day 5, independent of their status on Day 3. While Day 3 information is correlated with implantation potential, the effect is completely captured by the embryo's status on Day 5. For ovarian stimulation, the data provide no evidence that aggressive ovarian stimulation is detrimental to the oocyte, either in its ability to mature to metaphase II, to develop as an embryo, or, if transferred, to form a viable pregnancy (figures 3 and 5); moreover, the data constrain any of these effects to be small. Thus, a clinic should not be concerned about a potential trade-off between the quality and quantity of retrieved oocytes. Conversely, the data show that, at the clinic from which our data were derived, the ovarian stimulation drugs FSH and HMG are typically applied at saturating or slightly deleterious doses for follicular recruitment. Thus, the hormone dosage could presumably be slightly reduced here, to mitigate side effects such as ovarian hyper-stimulation syndrome or the high cost of stimulation drugs, without a large decrease in the number of retrieved oocytes.

The results we present here are inferred using data from only one clinic. As such, some aspects of our models reflect the practice at one particular clinic rather than general biology. The treatment portions of the models, such as FSH and HMG doses, transfer decisions and missingness, will change from clinic to clinic. Likewise, properties of the patient population, such as the joint distribution of patient age and BMI, will presumably change from clinic to clinic. By contrast, we suspect that the broader, structural relationships among hormones, oocytes and embryos reflect general biology that will be broadly applicable, although varying measurement standards across clinics may cause quantitative changes in these relationships. It will be interesting to test our models by looking at data both from other, non-clinical sources and from IVF clinics, especially from other countries.

More broadly, our results provide fundamental insights into the overall process of development. Unlike the sparseness that arises in theoretically motivated models simply as a way to manage complexity, the sparseness in our models is a property of the data. This sparseness implies that the biology itself is simple, consisting of modularized processes that are quantitatively siloed from one another. The simplicity is surprising given the many ways that oogenesis and embryogenesis are affected by diseases, including age-related infertility, endometriosis, sperm malfunction and polycystic ovary syndrome, all of which are present in our dataset. Overall, our results suggest that, despite their underlying complexities, oogenesis and embryogenesis are modular processes that result in simple, emergent behaviour and that a concise, quantitative understanding of the rest of development may be possible.

Data accessibility. Data are available at https://doi.org/10.6084/m9.figshare.15153525.

Authors' contributions. C.R. provided the data; B.L. performed the analysis; B.L., C.R. and D.N. identified relevant study questions; and B.L., C.R. and D.N. wrote the paper.

Competing interests. We declare we have no competing interests.

Funding. D.N., B.L. and C.R. acknowledge the National Institutes of Health for financial support (grant nos. R01HD092550-01 and R01HD104969-01).

Acknowledgements. We would like to acknowledge K. Walden, S. Kou, V. Manoharan and the Needleman lab for useful discussions. B.L. acknowledges the Harvard Quantitative Biology Initiative for support and useful discussions.

References

1. Zhu M, Zernicka-Goetz M. 2020 Principles of self-organization of the mammalian embryo. *Cell* **183**, 1467–1478. (doi:10.1016/j.cell.2020.11.003)
2. Wennekamp S, Mesecke S, Nédélec F, Hiiragi T. 2013 A self-organization framework for symmetry breaking in the mammalian embryo. *Nat. Rev. Mol. Cell Biol.* **14**, 452–459. (doi:10.1038/nrm3602)
3. Song Y, Shvartsman SY. 2020 Chemical embryology redux: metabolic control of development. *Trends Genet.* **36**, 577–586. (doi:10.1016/j.tig.2020.05.007)
4. Gross P, Kumar KV, Grill SW. 2017 How active mechanics and regulatory biochemistry combine to form patterns in development. *Annu. Rev. Biophys.* **46**, 337–356. (doi:10.1146/annurev-biophys-070816-033602)
5. Chan CJ, Heisenberg C-P, Hiiragi T. 2017 Coordination of morphogenesis and cell-fate specification in development. *Curr. Biol.* **27**, R1024–R1035. (doi:10.1016/j.cub.2017.07.010)
6. Fiorentino J, Torres-Padilla M-E, Scialdone A. 2020 Measuring and modeling single-cell heterogeneity and fate decision in mouse embryos. *Annu. Rev. Genet.* **54**, 167–187. (doi:10.1146/annurev-genet-021920-110200)
7. Clift D, Schuh M. 2013 Restarting life: fertilization and the transition from meiosis to mitosis. *Nat. Rev. Mol. Cell Biol.* **14**, 549–562. (doi:10.1038/nrm3643)
8. Shahbazi MN, Siggia ED, Zernicka-Goetz M. 2019 Self-organization of stem cells into embryos: a window on early mammalian development. *Science* **364**, 948–951. (doi:10.1126/science.aax0164)
9. Rossant J, Tam PP. 2009 Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136**, 701–713. (doi:10.1242/dev.017178)
10. White MD, Zenker J, Bissiere S, Plachta N. 2018 Instructions for assembling the early mammalian embryo. *Dev. Cell* **45**, 667–679. (doi:10.1016/j.devcel.2018.05.013)
11. Hassold T, Hunt P. 2001 To err (meiotically) is human: the genesis of human aneuploidy. *Nat. Rev. Genet.* **2**, 280–291. (doi:10.1038/35066065)
12. Niakan KK, Han J, Pedersen RA, Simon C, Pera RAR. 2012 Human pre-implantation embryo development. *Development* **139**, 829–841. (doi:10.1242/dev.060426)
13. Jaffe LA, Egbert JR. 2017 Regulation of mammalian oocyte meiosis by intercellular communication within the ovarian follicle. *Annu. Rev. Physiol.* **79**, 237–260. (doi:10.1146/annurev-physiol-022516-034102)
14. Sethna J. 2006 *Statistical mechanics: entropy, order parameters, and complexity*, vol. 14. Oxford, UK: Oxford University Press.
15. Bouchaud J-P, Potters M. 2003 *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge, UK: Cambridge University Press.
16. Bialek W. 2012 *Biophysics: searching for principles*. Princeton, NJ: Princeton University Press.
17. Needleman D, Dogic Z. 2017 Active matter at the interface between materials science and cell biology. *Nat. Rev. Mater.* **2**, 1. (doi:10.1038/natrevmats.2017.48)
18. Eckmann J-P, Tlusty T. 2021 Dimensional reduction in complex living systems: where, why, and how. *BioEssays* **43**, e2100062. (doi:10.1002/bies.202100062)
19. Halabi N, Rivoire O, Leibler S, Ranganathan R. 2009 Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**, 774–786. (doi:10.1016/j.cell.2009.07.038)
20. Taheri-Araghi S, Bradde S, Sauls JT, Hill NS, Levin PA, Paulsson J, Vergassola M, Jun S. 2015 Cell-size control and homeostasis in bacteria. *Curr. Biol.* **25**, 385–391. (doi:10.1016/j.cub.2014.12.009)
21. Sauls JT, Li D, Jun S. 2016 Adder and a coarse-grained approach to cell size homeostasis in bacteria. *Curr. Opin. Cell Biol.* **38**, 38–44. (doi:10.1016/j.ceb.2016.02.004)
22. Kohram M, Vashistha H, Leibler S, Xue B, Salman H. 2020 Bacterial growth control mechanisms inferred from multivariate statistical analysis of single-cell measurements. *Curr. Biol.* **31**, 955–964. (doi:10.1016/j.cub.2020.11.063)
23. Weenen C, Laven JS, von Bergh AR, Cranfield M, Groome NP, Visser JA, Kramer P, Fauser BC, Themmen AP. 2004 Anti-mullerian hormone expression pattern in the human ovary: potential implications for initial and cyclic follicle recruitment. *MHR: Basic Sci. Reprod. Med.* **10**, 77–83. (doi:10.1093/molehr/gah015)
24. La Marca A, Broekmans F, Volpe A, Fauser B, Macklon N. 2009 Anti-mullerian hormone (AMH): what do we still need to know? *Human Reprod.* **24**, 2264–2275. (doi:10.1093/humrep/dep210)
25. Durlinger A, Visser J, Themmen A. 2002 Regulation of ovarian function: the role of anti-Mullerian hormone. *Reproduction* **124**, 601–609.
26. La Marca A, Sighinolfi G, Radi D, Argento C, Baraldi E, Artenisio AC, Stabile G, Volpe A. 2010 Anti-mullerian hormone (AMH) as a predictive marker in assisted reproductive technology (ART). *Hum. Reprod. Update* **16**, 113–130. (doi:10.1093/humupd/dmp036)
27. Franasiak JM, Forman EJ, Hong KH, Werner MD, Upham KM, Treff NR, Scott Jr RT. 2014 The nature of aneuploidy with increasing age of the female partner: a review of 15 169 consecutive trophectoderm biopsies evaluated with comprehensive chromosomal screening. *Fertil. Steril.* **101**, 656–663.e1. (doi:10.1016/j.fertnstert.2013.11.004)
28. Talmor A, Dunphy B. 2015 Female obesity and infertility. *Best Pract. Res. Clin. Obstet. Gynaecol.* **29**, 498–506. (doi:10.1016/j.bpobgyn.2014.10.014)
29. Broughton DE, Moley KH. 2017 Obesity and female infertility: potential mediators of obesity's impact. *Fertil. Steril.* **107**, 840–847. (doi:10.1016/j.fertnstert.2017.01.017)
30. Elder K, Dale B. 2020 *In-vitro fertilization*. Cambridge, UK: Cambridge University Press.
31. MacKay DJ, Mac Kay DJ. 2003 *Information theory, inference and learning algorithms*. Cambridge, UK: Cambridge University Press.
32. MacKay DJ. 1992 Bayesian interpolation. *Neural Comput.* **4**, 415–447. (doi:10.1162/neco.1992.4.3.415)
33. Friedman N, Linial M, Nachman I, Pe'er D. 2000 Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **7**, 601–620. (doi:10.1089/106652700750050961)
34. Pearl J. 2009 *Causality*. Cambridge, UK: Cambridge University Press.
35. Bühlmann P, Kalisch M, Meier L. 2014 High-dimensional statistics with a view toward applications in biology. *Annu. Rev. Stat. Appl.* **1**, 255–278. (doi:10.1146/annurev-statistics-022513-115545)
36. Friedman N. 2004 Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805. (doi:10.1126/science.1094068)
37. Bishop CM. 2006 *Pattern recognition and machine learning*. Berlin, Germany: Springer.

38. Pellatt L, Rice S, Mason HD. 2010 Anti-Müllerian hormone and polycystic ovary syndrome: a mountain too high? *Reproduction* **139**, 825–833. (doi:10.1530/REP-09-0415)

39. Erickson JD. (1978 Down syndrome, paternal age, maternal age and birth order. *Ann. Hum. Genet.* **41**, 289–298. (doi:10.1111/j.1469-1809.1978.tb01896.x)

40. Morris JK, Mutton DE, Alberman E. 2002 Revised estimates of the maternal age specific live birth prevalence of Down's syndrome. *J. Med. Screen.* **9**, 2–6. (doi:10.1136/jms.9.1.2)

41. Gruhn JR *et al.* 2019 Chromosome errors in human eggs shape natural fertility over reproductive life span. *Science* **365**, 1466–1469. (doi:10.1126/science.aav7321)

42. Holubcová Z, Blayney M, Elder K, Schuh M. 2015 Error-prone chromosome-mediated spindle assembly favors chromosome segregation defects in human oocytes. *Science* **348**, 1143–1147. (doi:10.1126/science.aaa9529)

43. Webster A, Schuh M. 2017 Mechanisms of aneuploidy in human eggs. *Trends Cell Biol.* **27**, 55–68. (doi:10.1016/j.tcb.2016.09.002)

44. Musacchio A. 2015 The molecular biology of spindle assembly checkpoint signaling dynamics. *Curr. Biol.* **25**, R1002–R1018. (doi:10.1016/j.cub.2015.08.051)

45. Mihajlović AI, FitzHarris G. 2018 Segregating chromosomes in the mammalian oocyte. *Curr. Biol.* **28**, R895–R907. (doi:10.1016/j.cub.2018.06.057)

46. Holt JE, Jones KT. 2009 Control of homologous chromosome division in the mammalian oocyte. *Mol. Hum. Reprod.* **15**, 139–147. (doi:10.1093/molehr/gap007)

47. Howe K, FitzHarris G. 2013 Recent insights into spindle function in mammalian oocytes and early embryos. *Biol. Reprod.* **89**, 71. (doi:10.1095/biolreprod.113.112151)

48. Hugues J, Soussis J, Calderon I, Balasch J, Anderson R, Romeu A. 2005 Does the addition of recombinant LH in WHO group II anovulatory women over-responding to FSH treatment reduce the number of developing follicles? A dose-finding study. *Hum. Reprod.* **20**, 629–635. (doi:10.1093/humrep/deh682)

49. Wikland M, Bergh C, Borg K, Hillensjö T, Howles C, Knutsson A, Nilsson L, Wood M. 2001 A prospective, randomized comparison of two starting doses of recombinant FSH in combination with cetrorelix in women undergoing ovarian stimulation for IVF/ICSI. *Hum. Reprod.* **16**, 1676–1681. (doi:10.1093/humrep/16.8.1676)

50. Cha J, Sun X, Dey SK. 2012 Mechanisms of implantation: strategies for successful pregnancy. *Nat. Med.* **18**, 1754–1767. (doi:10.1038/nm.3012)

51. Little RJ, Rubin DB. 2019 *Statistical analysis with missing data*, vol. 793. New York, NY: John Wiley & Sons.

52. Racowsky C, Stern JE, Gibbons WE, Behr B, Pomeroy KO, Biggers JD. 2011 National collection of embryo morphology data into society for assisted reproductive technology clinic outcomes reporting system: associations among day 3 cell number, fragmentation and blastomere asymmetry, and live birth rate. *Fertil. Steril.* **95**, 1985–1989. (doi:10.1016/j.fertnstert.2011.02.009)

53. Vernon M, Stern JE, Ball GD, Wininger D, Mayer J, Racowsky C. 2011 Utility of the national embryo morphology data collection by the society for assisted reproductive technologies (SART): correlation between day-3 morphology grade and live-birth outcome. *Fertil. Steril.* **95**, 2761–2763. (doi:10.1016/j.fertnstert.2011.02.008)

54. Gardner DK, Balaban B. 2016 Assessment of human embryo development using morphological criteria in an era of time-lapse, algorithms and OMICS: is looking good still important? *MHR: Basic Sci. Reprod. Med.* **22**, 704–718. (doi:10.1093/molehr/gaw057)

55. Braude P, Bolton V, Moore S. 1988 Human gene expression first occurs between the four- and eight-cell stages of preimplantation development. *Nature* **332**, 459–461. (doi:10.1038/332459a0)

56. Rienzi L *et al.* 2015 No evidence of association between blastocyst aneuploidy and morphokinetic assessment in a selected population of poor-prognosis patients: a longitudinal cohort study. *Reprod. Biomed. Online* **30**, 57–66. (doi:10.1016/j.rbmo.2014.09.012)

57. Minasi MG *et al.* 2016 Correlation between aneuploidy, standard morphology evaluation and morphokinetic development in 1730 biopsied blastocysts: a consecutive case series study. *Hum. Reprod.* **31**, 2245–2254. (doi:10.1093/humrep/dew183)

58. Campbell A, Fishel S, Bowman N, Duffy S, Sedler M, Hickman CFL. 2013 Modelling a risk classification of aneuploidy in human embryos using non-invasive morphokinetics. *Reprod. Biomed. Online* **26**, 477–485. (doi:10.1016/j.rbmo.2013.02.006)

59. Kaser DJ, Racowsky C. 2014 Clinical outcomes following selection of human preimplantation embryos with time-lapse monitoring: a systematic review. *Hum. Reprod. Update* **20**, 617–631. (doi:10.1093/humupd/dmu023)

60. Hill MJ, Richter KS, Heitmann RJ, Graham JR, Tucker MJ, DeCherney AH, Browne PE, Levens ED. 2013 Trophectoderm grade predicts outcomes of single-blastocyst transfers. *Fertil. Steril.* **99**, 1283–1289.e1. (doi:10.1016/j.fertnstert.2012.12.003)

61. Ahlström A, Westin C, Reismer E, Wikland M, Hardarson T. 2011 Trophectoderm morphology: an important parameter for predicting live birth after single blastocyst transfer. *Hum. Reprod.* **26**, 3289–3296. (doi:10.1093/humrep/der325)

62. Macklon NS, Stouffer RL, Giudice LC, Fauser BC. 2006 The science behind 25 years of ovarian stimulation for *in vitro* fertilization. *Endocr. Rev.* **27**, 170–207. (doi:10.1210/er.2005-0015)

63. Akin E, Lacker HM. 1984 Ovulation control: the right number or nothing. *J. Math. Biol.* **20**, 113–132. (doi:10.1007/BF00285341)

64. Lacker H. 1981 Regulation of ovulation number in mammals. A follicle interaction law that controls maturation. *Biophys. J.* **35**, 433–454. (doi:10.1016/S0006-3495(81)84800-X)

65. Lacker HM, Akin E. 1988 How do the ovaries count? *Math. Biosci.* **90**, 305–332. (doi:10.1016/0025-5564(88)90072-7)

66. Leary C, Leese HJ, Sturmey RG. 2015 Human embryos from overweight and obese women display phenotypic and metabolic abnormalities. *Hum. Reprod.* **30**, 122–132. (doi:10.1093/humrep/deu276)

67. Van de Velde H, Cauffman G, Tournaye H, Devroey P, Liebaers I. 2008 The four blastomeres of a 4-cell stage human embryo are able to develop individually into blastocysts with inner cell mass and trophectoderm. *Hum. Reprod.* **23**, 1742–1747. (doi:10.1093/humrep/den190)

68. Willadsen S. 1980 The viability of early cleavage stages containing half the normal number of blastomeres in the sheep. *Reproduction* **59**, 357–362. (doi:10.1530/jrf.0.0590357)

69. Allen W, Pashen R. 1984 Production of monozygotic (identical) horse twins by embryo micromanipulation. *Reproduction* **71**, 607–613. (doi:10.1530/jrf.0.0710607)

70. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. 1999 From molecular to modular cell biology. *Nature* **402**, C47–C52. (doi:10.1038/35011540)

Inferring simple but precise quantitative models of human oocyte and early embryo development: Supplementary Materials

Brian D. Leahy[1,2], Catherine Racowsky[3,4], and Daniel Needleman[1,2,5]

[1]Dept. of Molecular and Cellular Biology, Harvard University, Cambridge MA
[2]SEAS, Harvard University, Cambridge, MA
[3]Brigham Women's Hospital, Boston, MA
[4]Harvard Medical School, Boston, MA
[5]Center for Computational Biology, Flatiron Institute, New York, NY

1 Inference and Regression Methods

1.1 Regressions

We perform nonlinear regression on the data by finding the polynomial model that maximises the Bayesian posterior model evidence. We do this in three steps, following the approach outlined in ref. [1]. First, we assume a class of probabilistic models for the data. For each model from the class, we then find the model parameters that maximise the posterior probability, given the model. Finally, we use these model regressions to evaluate which model has the highest posterior probability. In practice, this procedure depends on the choice of model parameterization and the choice of prior for the models and the parameters. We use two separate classes of probabilistic generative models when performing these regressions, one for regressing continuous data and one for regressing discrete data.

For continuous data, we use the following class of probabilistic models for the data. First, we scale both the dependent and regressor variables by their mean and standard deviation, such that they are mean 0 and variance 1. Next, we model the dependent variable as a function of the regressor variables, plus additive i.i.d. Gaussian noise of mean 0 and variance σ^2: $y_i = f(x_{ij}; \theta_\alpha) + \epsilon_i$, where θ_α are the parameters of the model. We fit both the model parameters and the noise standard deviation σ. We place a log-normal prior on the noise standard deviation, with mean parameter 0 and variance parameter 1. For x univariate, we choose Chebyshev polynomial series as the class of models. For x multivariate, we take $f(x_{ij}) = \sum_j C_j(x_j)$, where C_j is a Chebyshev series in each of the dependent variables. For both cases, we place priors on the coefficients as normal distributions with mean 0 and variance 1. As Chebyshev polynomials take values between 0 and 1 when the dependent variable is between 0 and 1, this model with this set of priors corresponds to assuming that $y(x)$ varies by roughly one standard deviation when x varies by roughly one standard deviation, with the stochastic and deterministic variation in the data being comparable to one another. Combined, the posterior probability of the model parameters given the model and the data is

$$\begin{aligned}\rho(\theta_\alpha, \sigma | y, \boldsymbol{x}; m) \propto & \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp\left(-\frac{1}{2}\left(\frac{y_i - f(x_{ij}, \theta_\alpha)}{\sigma}\right)^2\right) \times \\ & \prod_\alpha \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\theta_\alpha^2}{2}\right) \times \\ & \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\ln\sigma)^2}{2}\right)\end{aligned} \tag{1}$$

where m indexes the model, x_{ij}, y_i are the regressor and dependent variables for each datum, and f is a sum of Chebyshev series in each separate, scaled variable with coefficients θ_α. The first term corresponds to the likelihood, the second the priors on the model parameters, and the third the prior on the noise standard deviation. Including the terms $1/\sqrt{2\pi\sigma^2}$ in the likelihood ensures that the noise level fits to the correct value; the prior on the noise level has little effect on the model fitting.

For regressing discrete data, such as the missingness of a variable or the probability of a fetal heartbeat after transfer, we take a similar approach, but model the probability of success as a Bernoulli trial. We take the probability logit to be the same class of functions as before, *i.e.* the probability per trial is $p = 1/(1+\exp(-f(x_{ij}, \theta_\alpha)))$ where f is a sum of Chebyshev series in each separate, scaled variable.

Given a model from this class of models, we then fit the model's parameters from the posterior, by taking the model's maximum *a posteriori* parameters as point estimates and approximating the errors using a Laplace approximation on the posterior.

After the models have been fit, we perform model selection as outlined in ref [1]. We use a Laplace approximation to evaluate the probability of each model as

$$\rho(m|y, \boldsymbol{x}) = \rho(y|\boldsymbol{x}, \theta^*_\alpha, m) \times \rho(\theta^*_\alpha|m) \times \det\left(\mathbf{A}/2\pi\right)^{-1/2} \tag{2}$$

where $\rho(y|\boldsymbol{x}, \theta^*_\alpha, m)$ is the likelihood, $\rho(\theta^*_\alpha|m)$ is the prior distribution, θ^*_α are the maximum *a posteriori* parameters, and $\mathbf{A} = -\nabla\nabla\rho(\theta_\alpha|y, x; m)$ is the inverse of the parameters' covariance matrix. This procedure is related to the Bayesian Information Criterion [3]; the Bayes Information Criterion is an asymptotic approximation of this procedure in the limit of infinite samples. Qualitatively, the model selection combines two contrasting pieces of information: how well the model fits the data, and how fine-tuned the model's parameters need to be to describe the data. More complex models will in general fit the data better (*i.e.* have a higher likelihood), but will in general need more fine-tuning of their parameters (*i.e.* have a smaller determinant of the covariance matrix).

Since there are an infinite number of models in the space of models that we have assumed, we cannot fit all possible models and choose the one with the highest posterior probability. Instead, we search for the most probable model as follows. For univariate models, we exhaustively check all polynomial orders up to 8–10 and select the best model found, which is always less than the highest polynomial order checked. For multivariate models, we use the following heuristic search. Each point in the model space can be represented by a tuple of integers, with each integer representing the polynomial degree for the corresponding independent variable. To find the most probable model, we start by fitting the data to a quadratic in each variable. We then iteratively proceed by taking the best model found so far, increasing or decreasing one of the polynomial degrees individually, and re-fitting the model. If neither increasing nor decreasing any of the polynomial orders results in a better fit model, the algorithm terminates the search. For instance, when fitting MII to both Eggs and E2, the algorithm starts by fitting a second-order polynomial in both MII and E2, which can be represented by the point (2, 2).

The algorithm then checks the models (1, 2), corresponding to a model which is linear in Eggs and quadratic in E2, and finds that it is more probable than (2, 2). Next, the algorithm checks the point (0, 2). This is not more probable than (1, 2), so the algorithm proceeds to check the point (1, 1), which is linear in both Eggs and E2. This model is the most favored. The algorithm then checks the points (0, 1) and (1, 0), and finds that they are not favored over (1, 1). At this time, every point adjacent to the best point found so far has been checked, so the algorithm terminates, returning (1, 1) as the most likely model.

In practice, this process results in excellent fits to the data with well-defined model orders. We illustrate this with the regression of Eggs on AMH, shown in Fig. 1. Panel a shows the model probability and model likelihood as a function of the polynomial degree. The model likelihood continues to increase as the polynomial degree increases. In contrast, the model log-evidence attains a maximum at a fourth-degree polynomial; models with higher polynomial degrees require an unfavorable fine-tuning of parameters. Transforming these log-evidences to model probabilities gives a single model that is strongly preferred, with a probability of 97.4%, as shown in panel b. To check how well this process fits the data, we separate the raw data into bins and plot the bin averages as red circles in Fig. **??**c. The maximum-evidence model visually fits the data the best, as illustrated by panel c. That the maximum-evidence model fits the data well show that our choices for the space of models and the priors on model coefficients are sufficient to accurately fit the data. Additional comparisons of the maximum-evidence models can be seen in Main Text Fig. 1b,c, Main Text Fig. 5a, & SI Fig. 5a for regression with a Gaussian likelihood, and in Main Text Figures 5b-d & SI Fig. 6a,g for regression with a Poisson-binomial likelihood; in all cases the maximum-evidence model correctly fits the data. The regression process described above also performs accurately on generated data. For data generated with a linear model, this formalism identifies a linear model as the one that best fits the data. Likewise, for data generated with a quadratic model, the formalism identifies a quadratic model as best fitting the data. We implement these two as unit tests for our regression framework.

The results of the analyses in the main text are broadly robust to conditioning via this nonlinear regression vs a linear regression, although we do see some differences. One noticeable difference is in the graph in Fig. 3 in the main text. When we regress Eggs nonlinearly on AMH and HMG, we find that FSH is uncorrelated with Eggs, corresponding to a physiological regime where roughly all possible follicles are recruited by FSH. In contrast, conditioning using linear regression only gives a conditional correlation of FSH and Eggs that is statistically significantly negative, corresponding to a physiological regime where increasing FSH results in fewer recruited follicles. This apparent negative correlation arises because the mean of Eggs is increasing but concave-down as a function of AMH, whereas the mean of FSH is decreasing but concave-up as a function of AMH. As a result, linearly regressing on

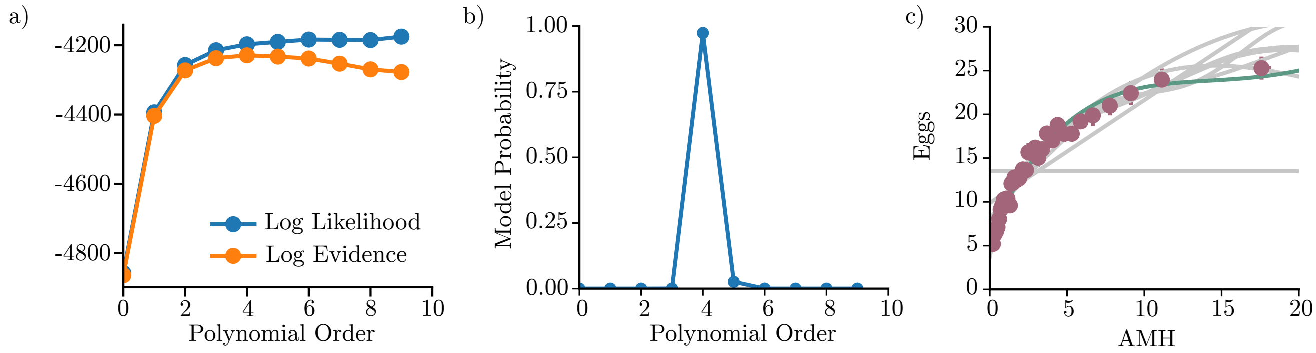


Figure 1: (a) The log-likelihood (blue) and log-evidence (orange) for polynomial models of degree 0-9, for regressing Eggs vs AMH. (b) The corresponding model probabilities. (c) The resulting regressions. Green line: maximum a-posteriori model, corresponding to a fourth-order polynomial. Gray lines: regressions corresponding to the models of other order. Red dots and error bars: mean ± standard error of Eggs vs AMH, after binning into 40 bins with equal counts per bin. The maximum a-posteriori model captures the nonlinearities in the data without overfitting. Compare to SI Fig. 4, which shows the maximum a-posteriori regression with the raw data for Eggs vs AMH.

AMH overestimates Eggs and underestimates FSH when AMH is low. Since the distribution of AMH is skewed, taking the correlation coefficient of these biased residuals then gives a negative correlation.

1.2 Estimating Fetal Heartbeat Probabilities

To estimate the probability of an embryo implanting, we use a forward (or generative) model and a Bayesian approach. We assume some class of models, each of which provides a probability of each embryo to develop sufficiently to provide a fetal heartbeat. We then use Bayes theorem to re-cast the probability distribution of fetal heartbeats, given both the model parameters and the number of embryos transferred, into a posterior distribution of the parameters given the measured number of fetal heartbeats and known number of embryos transferred, following the same approach outlined for the continuous regression. Here, we explain in the forward model for fetal heartbeats in more detail.

In each cycle, n embryos are transferred that result in h measured fetal heartbeats. We assume that each embryo has a probability p_e of leading to one fetal heartbeat, and a probability $1 - p_e$ of providing no fetal heartbeat, where the subscript e denotes the which embryo. For simplicity, we ignore the possibility of a single embryo forming monozygotic twins, which demographic data suggests should be a 1% correction. For the few cycles that do have more fetal heartbeats recorded than embryos transferred, we treat every embryo as successfully implanting. Then the probability $p(h|n, \{p_e\})$ is given by a Poisson-binomial distribution. For instance, if two embryos were transferred and one heartbeat was observed, then

$$p(h = 1|n = 2, \{p_e\}) = p_1(1 - p_2) + (1 - p_1)p_2 \quad , \tag{3}$$

where the subscript e indexes embryos and the subscript c indexes cycles.

The data is a collection of many treatment cycles, each with n_c embryos transferred and h_c fetal heartbeats measured. We treat each cycle as an independent event, and therefore the probability of the observed data is

$$p(\{h_c\}|\{n_c\}, \{p_{ec}\}) = \prod_c p(h_c|n_c, \{p_e\}_c) \tag{4}$$

where each $p(h_c|n_c, \{p_e\}_c)$ is a Poisson-binomial distribution.

To proceed further, we assume a functional form for the implantation probabilities p_{ec}. Each embryo has some associated parameters $\boldsymbol{x}_{ec}$ with it, such as the number of cells on day 3, the stage on day 5, the age of the woman, *etc.* We assume that the probability of forming a heartbeat depends on the parameters with some functional form f, which we parameterise with some set of parameters θ_α: $p_{ec} = p(\boldsymbol{x}_{ec}, \theta_\alpha)$. Substituting this into equation 4 allows writing of the probability of the forward model in terms of the (unknown) model parameters θ_α and the (known) embryo parameters $\boldsymbol{x}_{ec}$:

$$p(\{h_c\}|\{n_c\}, \theta_\alpha, \{\boldsymbol{x}_{ec}\}) = \prod_c p(h_c|n_c, \{p(\boldsymbol{x}_e; \theta_\alpha)\}_c) \tag{5}$$

We then fit the data with two types of models: smooth, low-dimensional models containing a few number of parameters, and with "model-independent" models that do not assume underlying smoothness of the functional form of $p(\boldsymbol{x}_{ec})$. For the low-dimensional models, we use the same approach as described earlier, parameterizing the logits with a Chebyshev series in the scaled input variables. For the model-independent approaches, we define $p(\boldsymbol{x};\theta_\alpha)$ as a piecewise-constant function over a series of intervals, by setting the parameters θ_α to the probability logits on each interval. For example, to calculate a model-independent measure of the probability of an embryo implanting given its day-5 stage, we assign one probability of implanting to all stage-1 embryos, another probability for stage-2 embryos, *etc.* For model-independent measures of the effect of age and BMI on embryo implantation potential, we bin the variables into separate intervals and assign the same probability to each interval; we choose the intervals to contain the same number of embryos (so the intervals are not of equal width). We place a prior on the logits that corresponds to a flat prior of the probabilities ($\rho(\ell) = \text{sech}^2(\ell/2)/4$, where ℓ is the logit). While this prior on the logit corresponds to a uniform prior for the fitted probability, this prior does have a maximum and therefore shifts the posterior maximum slightly from the maximum likelihood value. The low dimensional models should balance out variance and bias tradeoff; the model-independent models should be higher variance but do not make any assumptions about smoothness and therefore should have lower bias. These two models are plotted as the green lines and red dots, respectively, in Fig. 5 in the main text and in Fig. 6 in the SI.

After fitting the models, we perform Bayesian model selection, following the same formalism as described above for the regression models. To perform this model selection on day-3 vs day-5 variables, we only use the cycles that have data recorded for both day-3 and day-5. In principle the day-3-only cycles contain information about model selection, but to avoid subtleties due to other confounders we ignore these when performing model selection. In addition, we also limit the data to cycles with 4 or fewer embryo transfers (roughly 90% of the cycles with transfers), to avoid possible confounders when many embryos are transferred.

This approach for estimating the probability of fetal heartbeat assumes that embryos implant independently of one another in multiple transfers. We check this assumption by performing additional regressions with the number of transferred embryos as a variable. The regression results are consistent with multiple transfers systematically neither helping nor hurting the chance of an individual embryo to implant, as shown in SI Fig. 6g-h. While embryos that are part of a multiple transfer are much less likely to implant than those in a single transfer (panel g), this relationship disappears after controlling for the patient's age and the embryo's stage on day 5 (panel h).

1.3 Constructing DAGs

To construct the directed acyclic graph, we tailor our approach using prior knowledge of the data, rather than using more general algorithms that are agnostic to prior knowledge. One such agnostic algorithm for constructing directed acyclic graphs is the inductive-causation (IC) algorithm. The IC algorithm proceeds in three steps. In the first step, one draws a fully connected, undirected graph, then removes any edges A—B if there is any set of variables C such that A and B are conditionally independent given C. Once the undirected graph is constructed, then edges are oriented based on the presence of colliders [2]. This algorithm is guaranteed to produce a directed acyclic graph consistent with the data. We find that an IC algorithm does not perform well on the our dataset. The naive application of an IC algorithm constructs graphs that are physiologically nonsensical, presumably due to the finite statistical power we have in identifying whether two variables are conditionally independent. As an illustration from the data, consider the three variables Age, BMI, and E2. As discussed in the text, Age and BMI are weakly correlated, $\text{Corr}(\text{Age}, \text{BMI}) = 0.07$, and BMI and E2 are weakly negatively correlated, $\text{Corr}(\text{BMI}, \text{E2}) = -0.11$. However, Age and E2 appear uncorrelated, given nothing: $\text{Corr}(\text{Age}, \text{E2}) = -0.02$ ($P = 0.34$). An IC algorithm would suggest drawing the graph with the edges Age → BMI and E2 → BMI, suggesting that the patient's maximum estradiol concentration recorded during an IVF treatment is what determines whether or not she is overweight. This is especially absurd as some patients have multiple treatment cycles with different recorded E2 but the same BMI. Presumably, the apparent conditional independence between Age and E2 would disappear if much more data was collected. To avoid these types of difficulties, we enforce some prior knowledge in the structure of the directed acyclic graph, although we keep the prior knowledge to the minimum to minimise confirmation bias. Similar problems apply to using a LASSO approach to construcing the directed acyclic graphs.

2 Data description

The clinicians score embryos on day 3 and day 5 according to the following procedures:

- **Day 3 Cells**: The number of cells on day 3.
- **Day 3 Fragmentation**: The volume percent of the embryo occupied by fragments, scored as 0%, 1–10%, 11–25%, 26–50%, and >50%.
- **Day 3 Multinucleation**: Scored as 1 (at least 1 blastomere has more than 1 nucleus) or 0 (otherwise).
- **Day 3 Symmetry**: Scored as 1 (perfect symmetry), 2 (moderately asymmetric), or 3 (severely asymmetric).
- **Day 3 Vacuoles**: Scored as 0 (no vacuoles) or 1 (has vacuoles).
- **Day 3 Granularity**: Scored as 0 (not granular) or 1 (granular).
- **Day 5 Stage**: Scored from 1–9, as
  1. Degenerate or arrested; the embryo failed to develop to the morula stage.
  2. Morula, with incomplete compaction (less than 50% compacted).
  3. Morula, with more than 50% of the embryo compacted, but no blastocyst formation visible.
  4. Early blastocyst, where the blastocoele is less than half the volume of the embryo, with little to no expansion in the embryo's volume. The zona pellucida has not started thinning.
  5. Expanding Blastocyst, where the blastocoele occupies more than half the embryo's volume, with some expansion in the embryo's size and the zona pellucida starting to thin.
  6. Full blastocyst, where the blastocoele completely fills the embryo but the zona pellucida has not completely thinned.
  7. Expanded Blastocyst, where the blastocoele completely fills the embryo, which has fully expanded. The zona pellucida is very thin.
  8. Hatching Blastocyst, where the trophectoderm is starting to herniate through the zona pellucida.
  9. Hatched Blastocyst, where the blastocyst is completely hatched out of the zona pellucida.
- **Day 5 ICM**: Scored as a grade from 1–4, with
  1. ICM (inner cell mass) prominent and easily discernible, with many cells that are compacted and tightly adhered together.
  2. ICM discernible, but with fewer cells, and loosely adherent together.
  3. Very few cells visible, either compacted or loose. ICM cells be difficult to distinguish from trophectoderm.
  4. No cells visible in the ICM, or all cells are degenerate or necrotic.

  This is only scored for blastocysts (*i.e.* stage 5 and above).
- **Day 5 Trophectoderm**: Scored as a grade from 1–4, with
  1. A continuous layer of small, uniform, eye-shaped cells bordering the blastocoele.
  2. Fewer, larger cells that may not form a continuous layer.
  3. Sparse trophectoderm cells, which may be large.
  4. All trophectoderm cells are degenerate.

  This is only scored for blastocysts (*i.e.* stage 5 and above).

In addition, below we provide a list of all the variables describing both the ovarian stimulation and the development of the resulting oocytes, along with the percentage of cycles and oocytes for which those variables are recorded. In addition to the variables listed below, there are additional variables not listed that assist in linking oocytes to cycles.

Variable Name	% recorded, cycles	% included, oocytes	Notes
Cycle Code	100.0	100.0	An anonymous code specific to each cycle
Patient Code	100.0	100.0	An anonymous code specific to each patient
Cycle Type	100.0	100.0	fresh or frozen cycle
Date of Cycle Start	0.0	0.0	
Date of Retrieval	0.0	0.0	
GC	100.0	100.0	Whether a gestational carrier was used
Patient Age	100.0	100.0	The age of the woman seeking treatment
Producer Age	100.0	100.0	The age of the woman providing oocytes
Producer BMI	98.7	99.6	Body-mass index of the woman seeking treatment
Carrier BMI	88.3	78.7	Body-mass index of the woman providing oocytes
Producer Race	70.7	99.8	Race of the woman providing oocytes
Gravidity	100.0	100.0	Number of previous pregnancies
Parity	55.4	50.6	Number of previous deliveries
Prior SAB	55.4	50.6	Number of prior spontaneous abortions
Prior TAB	55.4	50.6	Number of prior therapeutic abortions
AMH	60.3	67.6	"AMH" in main text
Day3 FSH	62.8	89.6	Serum FSH on Day 3 of a natural cycle
Day3 E2	62.0	88.4	Serum estradiol on Day 3 of a natural cycle
Fresh Stim Protocol	68.1	99.7	Antagonist, agonist, *etc*
CET cycle type	31.8	0.2	if the frozen transfer was hormonally controlled
Prog Type	91.0	85.4	Progesterone
CET with lupron	31.1	0.2	Whether lupron was used in the frozen cycle
HMG	68.2	99.8	
FSH	68.2	99.8	
Baseline Endo	67.2	98.2	Endometrial thickness at fresh cycle start
E2 Max Day	68.2	99.8	The day that the max E2 was achieved
E2 D02	14.0	21.4	The E2 measurment on Day 02 *etc* of the cycle
E2 D03	0.1	0.1	
E2 D04	0.0	0.0	
E2 D05	0.1	0.2	
E2 D06	3.8	6.8	
E2 D07	62.9	91.3	
E2 D08	26.5	46.4	
E2 D09	50.0	76.2	
E2 D10	36.5	59.7	
E2 D11	42.0	65.4	
E2 D12	29.3	43.3	
E2 D13	21.7	29.0	
E2 D14	12.8	15.5	
E2 D15	7.8	8.9	
E2 D16	4.3	4.5	
E2 D17	2.4	2.3	
E2 D18	1.6	1.6	
E2 D19	1.0	1.0	
E2 D20	0.8	0.8	
E2 D21	0.4	0.5	
E2 D22	0.3	0.3	
E2 D23	0.2	0.3	
E2 D24	0.2	0.2	
E2 D25	0.1	0.2	
Ovulatory trigger	67.8	99.2	
Dose of hCG trigger(IU)	49.7	72.0	
Dose of lupron trigger	13.5	29.1	
Dose of ovidrel trigger	14.1	17.5	

Variable Name	% recorded, cycles	% included, oocytes	Notes
Day Ovulatory Trigger	68.1	99.8	
Trigger E2	67.4	98.6	E2 on the day of ovulatory trigger
USEndo	67.2	98.5	Endometrial thickness on day of ovulatory trigger
Num≥18	67.9	99.4	Number of follicles ≥18 mm in diameter
Num17	67.9	99.4	Number of follicles 17 mm in diameter
Num16	67.9	99.4	Number of follicles 16 mm in diameter
Num≥12≤16	67.9	99.4	Number of follicles between 12 and 16 mm
Prog Min	24.0	38.5	Min. progesterone level, fresh cycles
Prog Max	24.0	38.5	Max. progesterone level (ng/mL) in fresh cycles
E2 Max	68.2	99.8	"E2" in main text
Sum Eggs	68.2	99.8	"Eggs" in main text
Sum MII	68.2	99.8	"MII" in main text
Sperm Origin	68.7	99.7	ejaculate, epididymis, testicular
Sperm Source	68.6	99.6	patient or donor
Sperm Fresh or Frozen	68.7	99.7	
Sum Fert	68.2	99.8	Number of fertilized eggs
Transfer	100.0	100.0	Whether a transfer occured
Sum Transferred	100.0	100.0	Number of embryos transferred
Day of transfer	88.6	79.1	Measured from fertilization
ARTSite Result	99.9	100.0	Overall result of treatment
Days ET to HCG.1	88.0	78.2	Time from transfer to 1st hCG measurement
HCG.1	88.6	79.1	1st hCG measurement
Days ET to HCG.2	53.2	46.3	
HCG.2	51.6	44.9	
Days ET to HCG.3	44.6	37.2	
HCG.3	44.6	37.2	
Sacs Max	87.0	77.6	
FH>12wk	86.0	77.4	"FH" in main text
# Babies (>24wks)	82.9	75.1	
# Fetuses SR	45.6	49.3	number of fetuses selectively reduced
GA	29.3	25.3	Gestational age at delivery, weeks
Out fetus1	34.3	32.5	Outcome of fetus (live birth, stilbirth, etc)
Gender baby1	29.0	25.4	
Wt baby1	28.9	25.3	
Out fetus2	5.6	5.4	
Gender baby2	5.2	5.0	
Wt baby2	5.1	4.9	
Out fetus3	0.0	0.0	
Gender baby3	0.0	0.0	
Wt baby3	0.0	0.0	
DX Dim Ovar Rsrv	98.2	98.4	Patient diagnoses, boolean
DX Endometriosis	98.2	98.4	
DX MaleFactor	98.2	98.4	
DX Ovul dysf	98.2	98.4	
DX Tubal disease	98.2	98.4	
DX Unknown	98.2	98.4	
DX Uterine factor	99.4	99.8	
DX Other	98.2	98.4	

Variable Name	% recorded, cycles	% included, oocytes	Notes
Generic Code	—	0.0	An anonymous code specific to each oocyte
Donor Egg	—	100.0	Whether the egg was from a donor
Fresh Fate	—	100.0	Frozen, transferred, or discarded
Final Fate	—	100.0	Frozen, transferred, or discarded
Variable Name	% recorded, cycles	% included, oocytes	Notes
Fresh % O2	—	100.0	Culture oxygen percentage, pre-freeze
Post-thaw % O2	—	6.3	Culture oxygen percentage, post-freeze
Freeze Protocol	—	19.2	slow or vitrification
Number of Freezes	—	95.4	
Day Freeze 1	—	19.2	
Day Freeze 2	—	0.2	
Day Transfer	—	18.7	
Stage at Retrieval	—	100.0	Stage of the oocyte on retrieval
Fert Method	—	100.0	IVF or ICSI
Fert HPI	—	86.9	when the fertilization check was performed
Fert Status	—	85.9	fertilization status of the embryo
Fresh D3 HPI	—	56.0	when the Day 3 evaluation was performed
Fresh D3 Cnum	—	55.9	“Day 3 Cells” in the main text
Fresh D3 Fragmentation	—	55.9	
Fresh D3 Symmetry	—	55.9	
Fresh D3 PMD	—	55.3	whether cell membranes are clearly visible
Fresh D3 Granularity	—	55.9	
Fresh D3 Vacuoles	—	55.9	
Fresh D3 Multinucleation	—	55.9	
Fresh D5 HPI	—	42.4	When the Day 5 evaluation was performed
Fresh D5 Stage	—	42.3	“Day 5 Stage” in the main text
Fresh D5 ICM	—	15.2	Inner-cell mass grade
Fresh D5 TE	—	15.2	Trophectorderm grade
Fresh D5 St4 Quality	—	8.5	Separate quality grade for early blastocysts
Fresh D6 HPI	—	17.8	When the Day 6 evaluation was performed
Fresh D6 Stage	—	17.8	
Fresh D6 ICM	—	5.9	
Fresh D6 TE	—	5.9	
Fresh D6 St4 Quality	—	2.8	
Time 1st Frozen HPI	—	17.7	
Thaw D3 HPI	—	1.8	
Thaw D3 Cnum	—	1.8	
Thaw D3 Fr	—	1.7	
Thaw D3 Sym	—	1.7	
Thaw D3 Pmd	—	1.7	
Thaw D3 G	—	1.7	
Thaw D3 V	—	1.7	
Thaw D3 MNB	—	1.7	
Thaw D5 HPI	—	4.6	
Thaw 5 Stage	—	4.6	
Thaw D5 ICM	—	3.5	
Thaw D5 TE	—	3.5	
Thaw D5 st4 Qual	—	0.6	
Thaw D6 HPI	—	1.1	
Thaw D6 Stage	—	1.2	
Thaw D6 ICM	—	0.7	
Thaw D6 TE	—	0.7	
Thaw D6 st4 Qual	—	0.1	
% survival	—	6.3	Fraction of the embryo that survived freezing
AH	—	100.0	Whether assisted hatching was used
Biopsy	—	60.9	Whether the embryo was biopsied

Variable Name	% recorded, cycles	% included, oocytes	Notes
Biopsy Day	—	2.9	
Number Cells Biopsied	—	2.8	
TBR	—	2.9	If the embryo was thawed, biopsied, then refrozen
PGT-gender	—	1.9	Gender from genetic testing (PGT)
PGT-ploidy	—	2.1	
SGD	—	1.1	Presence of single-gene defects
Translocation	—	0.3	
HLA matching	—	0.1	A measurement of immunological compatibility
Xfer HPI	—	18.6	Time of transfer
Day0 Embryol	—	98.1	Anonymous ID of the fertilising embryologist
Fert Embryol	—	86.8	ID of the embryologist checking fertilisation
D3 FrEval Embryol	—	91.9	ID of Day 3 fresh evaluating embryologist
D3ThEval Embryol	—	44.0	ID of Day 3 thawed evaluating embryologist
D5 FrEval Embryol	—	11.5	ID of Day 5 fresh evaluating embryologist
D5 ThEval Embryol	—	6.1	ID of Day 5 thawed evaluating embryologist
D6 FrEval Embryol	—	6.1	ID of Day 6 fresh evaluating embryologist
D6 ThEval Embryol	—	0.3	ID of Day 6 thawed evaluating embryologist
PGD Embryol	—	3.8	ID of embryologist performing genetic testing
Transfer Embryol	—	18.1	ID of embryologist performing transfer
Transfer MD	—	18.0	ID of the doctor performing transfer
PGD Result	—	0.0	

3 Structural Equations corresponding to Models

3.1 Structural models for Ovarian Stimulation and Pre-implantation Development

As explained in Sec. 1.1, we assume that the data is normally distributed, with a constant standard deviation and a mean that depends on the dependent variables. In reality, the data show strong evidences of heteroskedasticity and non-normality, cf. Fig. 3 in the main text. Nevertheless these structural models give a reasonably accurate description of the data.

The structural models for ovarian stimulation are calculated using a complete-case basis, since the missingness patterns appears to be correlated with changing clinical practices over time and not with any of the variables. There are 3422 cycles in the train set which have all of AMH, Eggs, MII, and E2 recorded, but 9 of these cycles are missing at least one of Age, BMI, FSH, or HMG. These missing 9 cycles are responsible for the small difference in the MII structural equations between that for the 4-element model and the 8-element model. Differences between other equations correspond to the effects of including other variables.

The structural models for pre-implantation development are calculated using all the data available for each equation; as such, different embryos and cycles are used to model the effect of Age on Day 3 Cells than are used to model Age on Day 5 Stage. In all these equations, Age is measured in years, BMI in kg / m^2, E2 is measured in pg / mL, and AMH is measured in international units (IU). The equations for FSH and HMG describe dosage in ampules; one ampule of FSH contains 150 IU, whereas one ampule of HMG contains 75 IU each of LH and FSH. The Day 3 and Day 5 variables are measured as described in Sec. 2.

3.1.1 Structural models for Ovarian Stimulation, Main Text Fig. 1

$$\begin{aligned} \text{AMH} =& N(\mu, \sigma^2), \text{with} \\ \mu =& 3.100 \\ \sigma =& 3.706 \end{aligned} \tag{6}$$

$$\begin{aligned} \text{Eggs} =& N(\mu, \sigma^2), \text{with} \\ \mu =& 6.078 + \\ & 3.938 \times \text{AMH} - 0.325 \times \text{AMH}^2 + 0.012 \times \text{AMH}^3 - 1.38 \times 10^{-4} \times \text{AMH}^4 \\ \sigma =& 7.359 \end{aligned} \tag{7}$$

$$\begin{aligned} \text{E2} =& N(\mu, \sigma^2), \text{with} \\ \mu =& 9.76 \times 10^2 + \\ & 96.252 \times \text{Eggs} - 0.957 \times \text{Eggs}^2 \\ \sigma =& 8.89 \times 10^2 \end{aligned} \tag{8}$$

$$\begin{aligned} \text{MII} =& N(\mu, \sigma^2), \text{with} \\ \mu =& 0.133 + \\ & 0.695 \times \text{Eggs} + \\ & 2.66 \times 10^{-4} \times \text{E2} \\ \sigma =& 2.929 \end{aligned} \tag{9}$$

3.1.2 Structural models for Ovarian Stimulation, Main Text Fig. 3

$$\begin{aligned} \text{Age} =& N(\mu, \sigma^2), \text{with} \\ \mu =& 36.563 \\ \sigma =& 4.357 \end{aligned} \tag{10}$$

$$\begin{aligned}
\mathrm{BMI} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 22.087 + \\
& 0.110 \times \mathrm{Age} \\
\sigma =& 6.570
\end{aligned} \tag{11}$$

$$\begin{aligned}
\mathrm{AMH} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 12.066 + \\
& -0.245 \times \mathrm{Age} \\
\sigma =& 3.543
\end{aligned} \tag{12}$$

$$\begin{aligned}
\mathrm{FSH} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 11.669 + \\
& 0.653 \times \mathrm{Age} + \\
& 0.172 \times \mathrm{BMI} + \\
& -7.928 \times \mathrm{AMH} + 0.780 \times \mathrm{AMH}^2 - 0.029 \times \mathrm{AMH}^3 + 3.49 \times 10^{-4} \times \mathrm{AMH}^4 \\
\sigma =& 15.516
\end{aligned} \tag{13}$$

$$\begin{aligned}
\mathrm{HMG} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 56.312 + \\
& -2.893 \times \mathrm{Age} + 0.054 \times \mathrm{Age}^2 + \\
& -20.019 \times \mathrm{AMH} + 4.181 \times \mathrm{AMH}^2 - 0.389 \times \mathrm{AMH}^3 + 0.018 \times \mathrm{AMH}^4 + \\
& -3.73 \times 10^{-4} \times \mathrm{AMH}^5 + 2.99 \times 10^{-6} \times \mathrm{AMH}^6 + \\
& 0.627 \times \mathrm{FSH} - 5.73 \times 10^{-3} \times \mathrm{FSH}^2 \\
\sigma =& 14.878
\end{aligned} \tag{14}$$

$$\begin{aligned}
\mathrm{Eggs} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 8.039 + \\
& 2.451 \times \mathrm{AMH} - 0.114 \times \mathrm{AMH}^2 + 1.59 \times 10^{-3} \times \mathrm{AMH}^3 + \\
& -0.066 \times \mathrm{HMG} \\
\sigma =& 7.308
\end{aligned} \tag{15}$$

$$\begin{aligned}
\mathrm{E2} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 1.14 \times 10^3 + \\
& 26.308 \times \mathrm{Age} + \\
& -14.213 \times \mathrm{BMI} + \\
& -4.655 \times \mathrm{FSH} + \\
& 13.637 \times \mathrm{HMG} - 0.169 \times \mathrm{HMG}^2 + \\
& 1.04 \times 10^2 \times \mathrm{Eggs} - 1.042 \times \mathrm{Eggs}^2 \\
\sigma =& 8.65 \times 10^2
\end{aligned} \tag{16}$$

$$\begin{aligned}
\mathrm{MII} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 0.132 + \\
& 0.695 \times \mathrm{Eggs} + \\
& 2.64 \times 10^{-4} \times \mathrm{E2} \\
\sigma =& 2.932
\end{aligned} \tag{17}$$

3.1.3 Structural models for Pre-Implantation Development, Main Tex Fig. 5

$$\begin{aligned}
\text{BMI} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 21.557+ \\
& 0.123 \times \text{Age} \\
\sigma =& 6.428
\end{aligned} \tag{18}$$

$$\begin{aligned}
\text{MII} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 42.519+ \\
& -1.494 \times \text{Age} + 0.016 \times \text{Age}^2 \\
\sigma =& 6.571
\end{aligned} \tag{19}$$

$$\begin{aligned}
\text{Day 3 Cells} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 8.783+ \\
& -0.038 \times \text{Age} \\
\sigma =& 2.429
\end{aligned} \tag{20}$$

$$\begin{aligned}
\text{Day 5 Stage} =& N(\mu, \sigma^2), \text{with} \\
\mu =& 3.642+ \\
& 0.063 \times \text{Age} - 1.58 \times 10^{-3} \times \text{Age}^2+ \\
& 0.847 \times \text{Day 3 Cells} - 0.031 \times \text{Day 3 Cells}^2 \\
\sigma =& 1.828
\end{aligned} \tag{21}$$

$$\begin{aligned}
P(\text{Fetal Heartbeat}) =& (1 + e^{-z})^{-1}, \text{with} \\
z =& -8.704+ \\
& 0.543 \times \text{Age} - 9.18 \times 10^{-3} \times \text{Age}^2+ \\
& 0.292 \times \text{Day 5 Stage}
\end{aligned} \tag{22}$$

We present the missingness structural equations here for completeness. Empirically, the distribution of whether an embryo is transferred (*i.e.* missingness for fetal heartbeat) differs for embryos transferred on day 3 versus those on day 5; we present both here as separate equations.

$$\begin{aligned}
P(\text{D5 Rec.}) =& (1 + e^{-z})^{-1}, \text{with} \\
z =& -3.588+ \\
& 0.923 \times \text{MII} - 0.054 \times \text{MII}^2 + 1.35 \times 10^{-3} \times \text{MII}^3 - 1.12 \times 10^{-5} \times \text{MII}^4+ \\
& -2.587 \times \text{Age} + 0.089 \times \text{Age}^2 - 9.95 \times 10^{-4} \times \text{Age}^3+ \\
& 4.369 \times \text{Day 3 Cells} - 1.222 \times \text{Day 3 Cells}^2 + 0.147 \times \text{Day 3 Cells}^3- \\
& 7.91 \times 10^{-3} \times \text{Day 3 Cells}^4 + 1.57 \times 10^{-4} \times \text{Day 3 Cells}^5+ \\
& -0.129 \times \text{BMI} + 1.99 \times 10^{-3} \times \text{BMI}^2
\end{aligned} \tag{23}$$

$$\begin{aligned}
P(\text{Trans.}|\text{D5 Rec.}) =& (1 + e^{-z})^{-1}, \text{with} \\
z =& -3.776+ \\
& 0.129 \times \text{MII} - 9.82 \times 10^{-3} \times \text{MII}^2 + 1.23 \times 10^{-4} \times \text{MII}^3+ \\
& 1.604 \times \text{Age} - 0.056 \times \text{Age}^2 + 6.42 \times 10^{-4} \times \text{Age}^3+ \\
& 2.300 \times \text{Day 3 Cells} - 0.223 \times \text{Day 3 Cells}^2 + 6.60 \times 10^{-3} \times \text{Day 3 Cells}^3+ \\
& 1.090 \times \text{Day 5 Stage} - 0.029 \times \text{Day 5 Stage}^2
\end{aligned} \tag{24}$$

$$
\begin{aligned}
P(\text{Trans.}|\text{D5 Miss.}) =&(1+e^{-z})^{-1}, \text{with} \\
z =&6.151+ \\
&-0.760 \times \text{MII} + 0.028 \times \text{MII}^2 - 2.92 \times 10^{-4} \times \text{MII}^3 + \\
&1.977 \times \text{Age} - 0.076 \times \text{Age}^2 + 9.52 \times 10^{-4} \times \text{Age}^3 + \\
&1.568 \times \text{Day 3 Cells} - 0.082 \times \text{Day 3 Cells}^2
\end{aligned} \tag{25}
$$

4 Supporting Data & Graphs for claims in the main text

4.1 Cross-Validation Results

The rank plot in Figure 3b in the main text shows the measured P-values for 99 conditional correlations among the 8 variables corresponding to the 99 conditional independencies predicted by the model in Figure 3a in the main text. To create this plot, we measure the conditional correlation and associated P-value for each conditional correlation. We then sort and plot the P-values. We do this process separately for the train (green) and test (red) data. The table at the end of this section lists all measured P-values on the test and train sets; note that the conditional correlations with suspiciously low P-values on the test set typically do not have low P-values on the train set, and vice versa.

Next, we check if the 99 P-values measured from the data are consistent with what one would expect if the model is true. If the model is correct, then all of these conditional correlations should be consistent with zero, and none of the measured P-values should be statistically significant. There are two complications here. First, since there are 99 separate tests, we must correct for multiple hypothesis testing. Second, since these 99 conditional correlations are measured from the same 8 variables, these 99 tests are not independent – for example, Corr(MII, Age | E2, Eggs) and Corr(MII, Age | E2, Eggs, AMH) test different independencies but involve the same variables MII and Age. We account for these complications by comparing the P-values from the measured conditional correlations in the data with those from a distribution of 3000 datasets simulated according to the proposed model. If the proposed model is correct, then the 99 conditional correlations and P-values measured from the dataset should be consistent with being drawn from the distribution of simulated P-values.

We simulate 3000 datasets with the following procedure. First, we randomly sample Age with replacement. We then generate the BMI by combining the structural equations from Main Text Figure 3a and SI Section 3 with the randomly-sampled (with replacement) residuals from the fit of BMI to the train data. We then continue this process by proceeding down the graph in Main Text Figure 3a to calculate AMH, then FSH, HMG, *etc.*. We then repeat this process for a total 3000 times, to produce 3000 simulated datasets.

For each simulated dataset, we then calculate the 99 conditional correlations by performing the necessary regressions with the same polynomial degree used for the real data and calculating the correlation between the residuals. These simulated P-value rank plots give an estimate of what would be expected if the model proposed in the main text is true. In principle, the distribution of P-value ranks for those generated according to the train data procedure, where each dataset containg 3,413 cycles and the conditional correlations are evaluated by regressing on that dataset, differs from the rank distribution generated according to the test data procedure, where each dataset contains 1,497 cycles and the conditional correlations are evaulated using regressions on the train data. In practice, the differences between these distributions are not visible; as such, Main Text Figure 3 just shows the distribution for the simulated train data. For comparison, we also generate rank plots according to linear Gaussian fully-connected models (Main Text Figure 3c). For this, we first generate 3,000 datasets according to a random Gaussian models, each with a randomly-drawn covariance matrix $C = UU^T$, where U_{ij} is a Gaussian random variable with mean 0 and variance 1. We then calculate the 99 conditional correlations for each of those 3,000 datasets, performing the regressions with the same polynomial degree used for the real data and the previously simulated data.

We quantify the similarity between the expected and measured rank plots in Main Text Figure 3b with two statistics. First, we examine the maximum distance between the measured P-value ranks and the median expected from the simulation (*i.e.* the maximum vertical separation between the black line and the red or green lines in the figure), similar to a Kolmogorov-Smirnov test. The maximum distance for the P-values measured on the train set is 0.33; 0.09 of the simulated rank curves have this maximum distance or greater. The maximum distance for the P-values measured on the test set is 0.24; 0.31 of the simulated rank curves have this maximum distance or greater. This first statistic suggests that the proposed model broadly agrees with the data. Second, we examine the minimum P-value for the 99 conditional independencies. The minimum P-value measured on the training set is 2×10^{-4} (corresponding to $\mathrm{Corr}(\mathrm{AMH}, \mathrm{E2} \mid \mathrm{Age}, \mathrm{BMI}, \mathrm{FSH}, \mathrm{HMG}, \mathrm{Eggs}, \mathrm{MII})$); only 0.004 of the simulated rank curves have a P-value this low. The minimum P-value measured on the test set is 7×10^{-5} (corresponding to $\mathrm{Corr}(\mathrm{Age}, \mathrm{Eggs} \mid \mathrm{AMH}, \mathrm{FSH}, \mathrm{HMG})$); only 0.001 of the simulated rank curves have a P-value this low. Thus, this second statistic weakly suggests that some features are missing from the model. Combined, this analysis shows that the model in Main Text Fig. 3 is broadly consistent with the data, although with some evidence for additional, small physiological effects missing from the model.

We also meaure how much the variance of the residuals changes when including all possible parameters, for both the ovarian stimulation graph and the development graph. To do this, we fit the train data assuming a completely connected graph. We nonlinearly regress each variable on all the upstream variables, following the procedure outlined in Sec. 1.1 but forcing each term to enter in at least linearly. We then measure the variance of the residuals on the test set, and compare the variance to the residuals on the test set using the model proposed in the main text. Of the

8 ovarian stimulation variables shown in Fig. 3 in the main text, 3 remain unchanged on changing to a fully connected model: Age is the first variable in both graphs and has no edges pointing into it, BMI is the second variable in both graphs and therefore only has Age pointing into it, and FSH has all upstream variables pointing into it in the fully connected model. The variance change for the remaining 5 variables is shown in the table below. Using a complete model worsens the fits for all but Eggs. This worsening is presumably due to the increased variance in the regression estimate when including additional variables. For pre-implantation development, the variance of the Day 3 Cells residuals decreases when including MII and BMI, whereas the variance of the Day 5 Stage residuals increases.

Name	Test Set Var., proposed model	Test Set Var., complete model	Percent change
AMH	15.18	15.20	+0.153%
HMG	251.27	251.45	+0.071%
Eggs	51.68	51.37	-0.600%
E2	883,606	888,766	+0.584%
MII	8.98	8.99	+0.182%
Day 3 Cells	5.78	5.77	-0.184%
Day 5 Stage	3.40	3.40	+0.034%

To check the results for fetal heartbeat, we use a likelihood test. We fit two models on the train set, the model proposed in Main Text Fig. 5 and a complete model that includes additional parameters as described above. We then fix the model parameters to their maximum *a posteriori* values and calculate the likelihood of the models on the test set. Since the models are fit on the train set and evaluated on the test set, the ratio of the likelihood corresponds to a Bayesian odds ratio of the two models; the table below reports this as a probability of the model in the text being correct. The first row compares the model Age, Day 5 Stage → FH against the model Age, BMI, MII, Day 3 Cells, Day 5 Stage → FH; the second row compares the model Age, Day 5 Troph, Day 5 Stage → FH against Age, Day 3 Cells, Day 3 Frag, Day 3 Granularity, Day 3 Multinucleation, Day 3 Symmetry, Day 3 Vacuoles, Day 5 Stage, Day 5 ICM, Day 5 Trophectoderm → FH. The two sets of likelihood are not directly comparable, as not all Day 5 transfers have the trophectoderm and ICM grade recorded (these are only recorded for developed blastocysts of stage 5 or higher).

Name	Log Likelihood, proposed	Log Likelihood, complete	P
Including BMI, MII, and Day 3 Cells	-479.96	-478.14	0.14
Including all Day 3, Day 5 grades	-415.51	-415.61	0.53

Name	Corr., Train	P, Train	Corr., Test	P, Test
Corr(Age, Eggs | AMH, FSH, HMG)	-0.040	0.020	-0.102	7×10^{-5}
Corr(Age, Eggs | AMH, BMI, FSH, HMG)	-0.040	0.020	-0.102	7×10^{-5}
Corr(Age, Eggs | AMH, BMI, HMG)	-0.045	0.009	-0.090	5×10^{-4}
Corr(Age, Eggs | AMH, HMG)	-0.045	0.009	-0.090	5×10^{-4}
Corr(Age, MII | AMH, E2, Eggs)	0.012	0.467	-0.055	0.034
Corr(Age, MII | AMH, BMI, E2, Eggs)	0.013	0.453	-0.054	0.038
Corr(Age, MII | AMH, E2, FSH, Eggs)	0.009	0.615	-0.053	0.041
Corr(Age, MII | AMH, BMI, E2, FSH, Eggs)	0.009	0.615	-0.053	0.041
Corr(Age, MII | E2, Eggs)	0.016	0.351	-0.053	0.042
Corr(Age, MII | BMI, E2, Eggs)	0.016	0.338	-0.051	0.047
Corr(Age, MII | BMI, E2, FSH, Eggs)	0.010	0.544	-0.051	0.049
Corr(Age, MII | E2, FSH, Eggs)	0.010	0.544	-0.051	0.049
Corr(AMH, MII | Age, BMI, FSH, HMG, Eggs)	-0.030	0.083	-0.049	0.057
Corr(AMH, MII | Age, E2, FSH, HMG, Eggs)	-0.033	0.054	-0.047	0.066
Corr(AMH, MII | Age, BMI, E2, FSH, HMG, Eggs)	-0.033	0.054	-0.047	0.066
Corr(AMH, MII | BMI, E2, FSH, HMG, Eggs)	-0.035	0.043	-0.047	0.071
Corr(AMH, MII | E2, FSH, HMG, Eggs)	-0.035	0.043	-0.047	0.071
Corr(FSH, MII | E2, HMG, Eggs)	0.012	0.498	-0.040	0.126
Corr(FSH, MII | BMI, E2, HMG, Eggs)	0.012	0.483	-0.039	0.130
Corr(Age, MII | AMH, E2, FSH, HMG, Eggs)	0.029	0.093	-0.039	0.131
Corr(Age, MII | BMI, E2, FSH, HMG, Eggs)	0.029	0.093	-0.039	0.131
Corr(Age, MII | E2, FSH, HMG, Eggs)	0.029	0.093	-0.039	0.131
Corr(Age, MII | AMH, BMI, E2, FSH, HMG, Eggs)	0.029	0.093	-0.039	0.131
Corr(HMG, MII | E2, Eggs)	-0.039	0.022	-0.037	0.155
Corr(Age, MII | AMH, E2, HMG, Eggs)	0.029	0.086	-0.036	0.164
Corr(Age, MII | AMH, BMI, E2, HMG, Eggs)	0.029	0.085	-0.036	0.169
Corr(HMG, MII | BMI, E2, Eggs)	-0.039	0.022	-0.035	0.175
Corr(FSH, MII | Age, E2, HMG, Eggs)	0.003	0.874	-0.034	0.183
Corr(FSH, MII | Age, BMI, E2, HMG, Eggs)	0.003	0.854	-0.034	0.188
Corr(AMH, MII | Age, BMI, E2, HMG, Eggs)	-0.034	0.047	-0.034	0.191
Corr(AMH, MII | Age, E2, HMG, Eggs)	-0.034	0.047	-0.034	0.191
Corr(Age, MII | E2, HMG, Eggs)	0.033	0.055	-0.033	0.206
Corr(Age, MII | BMI, E2, HMG, Eggs)	0.033	0.054	-0.032	0.212
Corr(HMG, MII | AMH, E2, Eggs)	-0.037	0.031	-0.031	0.234
Corr(HMG, MII | AMH, BMI, E2, Eggs)	-0.037	0.031	-0.029	0.257
Corr(HMG, MII | E2, FSH, Eggs)	-0.047	0.006	-0.029	0.260
Corr(HMG, MII | BMI, E2, FSH, Eggs)	-0.047	0.006	-0.028	0.272
Corr(FSH, MII | E2, Eggs)	0.009	0.615	-0.027	0.290
Corr(AMH, MII | Age, BMI, E2, FSH, Eggs)	-0.017	0.328	-0.027	0.296
Corr(AMH, MII | Age, E2, FSH, Eggs)	-0.017	0.328	-0.027	0.296
Corr(BMI, Eggs | AMH, HMG)	-0.012	0.466	-0.027	0.296
Corr(AMH, MII | Age, BMI, E2, Eggs)	-0.021	0.225	-0.026	0.306
Corr(AMH, MII | Age, E2, Eggs)	-0.021	0.225	-0.026	0.306
Corr(HMG, MII | AMH, E2, FSH, Eggs)	-0.042	0.013	-0.026	0.306
Corr(HMG, MII | AMH, BMI, E2, FSH, Eggs)	-0.042	0.013	-0.026	0.322
Corr(FSH, MII | BMI, E2, Eggs)	0.009	0.586	-0.025	0.325
Corr(AMH, E2 | Age, BMI, FSH, HMG, Eggs)	0.061	3×10^{-4}	-0.024	0.355
Corr(BMI, Eggs | Age, AMH, HMG)	-0.011	0.508	-0.024	0.355
Corr(AMH, MII | BMI, E2, HMG, Eggs)	-0.032	0.066	-0.023	0.364
Corr(AMH, MII | E2, HMG, Eggs)	-0.032	0.066	-0.023	0.364
Corr(BMI, Eggs | AMH, FSH, HMG)	-0.011	0.540	-0.023	0.368
Corr(BMI, Eggs | Age, AMH, FSH)	-0.011	0.530	-0.023	0.370
Corr(BMI, Eggs | Age, AMH, FSH, HMG)	-0.011	0.539	-0.022	0.389

Name	Corr., Train	P, Train	Corr., Test	P, Test
Corr(AMH, MII | BMI, E2, FSH, Eggs)	-0.018	0.292	-0.022	0.397
Corr(AMH, MII | E2, FSH, Eggs)	-0.018	0.292	-0.022	0.397
Corr(FSH, MII | AMH, E2, Eggs)	0.021	0.226	-0.022	0.403
Corr(HMG, MII | Age, BMI, E2, Eggs)	-0.044	0.009	-0.021	0.424
Corr(HMG, MII | Age, BMI, E2, FSH, Eggs)	-0.049	0.004	-0.021	0.425
Corr(HMG, MII | Age, E2, FSH, Eggs)	-0.049	0.004	-0.021	0.425
Corr(AMH, E2 | Age, BMI, FSH, HMG, Eggs, MII)	0.064	2×10^{-4}	-0.020	0.429
Corr(FSH, MII | AMH, BMI, E2, Eggs)	0.021	0.220	-0.020	0.431
Corr(HMG, MII | Age, E2, Eggs)	-0.042	0.014	-0.018	0.475
Corr(BMI, MII | AMH, E2, Eggs)	-0.005	0.785	-0.018	0.487
Corr(BMI, MII | E2, Eggs)	-0.005	0.785	-0.018	0.487
Corr(HMG, MII | Age, AMH, E2, Eggs)	-0.042	0.015	-0.018	0.494
Corr(HMG, MII | Age, AMH, BMI, E2, FSH, Eggs)	-0.045	0.008	-0.017	0.502
Corr(HMG, MII | Age, AMH, E2, FSH, Eggs)	-0.045	0.008	-0.017	0.502
Corr(HMG, MII | Age, AMH, BMI, E2, Eggs)	-0.042	0.015	-0.017	0.515
Corr(BMI, MII | AMH, E2, FSH, Eggs)	-0.006	0.737	-0.017	0.516
Corr(BMI, MII | E2, FSH, Eggs)	-0.006	0.737	-0.017	0.516
Corr(BMI, MII | AMH, E2, HMG, Eggs)	-6×10^{-4}	0.971	-0.016	0.535
Corr(BMI, MII | E2, HMG, Eggs)	-6×10^{-4}	0.971	-0.016	0.535
Corr(BMI, MII | E2, FSH, HMG, Eggs)	-0.003	0.846	-0.015	0.554
Corr(BMI, MII | Age, E2, FSH, HMG, Eggs)	-0.003	0.846	-0.015	0.554
Corr(BMI, MII | AMH, E2, FSH, HMG, Eggs)	-0.003	0.846	-0.015	0.554
Corr(FSH, Eggs | AMH, HMG)	-0.040	0.019	-0.015	0.571
Corr(BMI, MII | Age, AMH, E2, Eggs)	-0.005	0.751	-0.014	0.577
Corr(BMI, MII | Age, E2, Eggs)	-0.005	0.751	-0.014	0.577
Corr(BMI, MII | Age, AMH, E2, FSH, HMG, Eggs)	-0.006	0.709	-0.014	0.584
Corr(BMI, MII | Age, AMH, E2, FSH, Eggs)	-0.006	0.709	-0.014	0.584
Corr(BMI, MII | Age, E2, FSH, Eggs)	-0.006	0.709	-0.014	0.584
Corr(AMH, MII | BMI, E2, Eggs)	-0.022	0.197	-0.013	0.604
Corr(AMH, MII | E2, Eggs)	-0.022	0.197	-0.013	0.604
Corr(BMI, MII | Age, AMH, E2, HMG, Eggs)	-0.002	0.894	-0.013	0.608
Corr(BMI, MII | Age, E2, HMG, Eggs)	-0.002	0.894	-0.013	0.608
Corr(FSH, MII | AMH, E2, HMG, Eggs)	0.008	0.655	-0.013	0.619
Corr(FSH, Eggs | AMH, BMI, HMG)	-0.045	0.009	-0.013	0.622
Corr(FSH, MII | Age, AMH, E2, Eggs)	0.018	0.296	-0.013	0.625
Corr(FSH, MII | Age, AMH, BMI, E2, Eggs)	0.018	0.288	-0.012	0.650
Corr(FSH, Eggs | Age, AMH, BMI, HMG)	-0.026	0.126	0.012	0.651
Corr(FSH, Eggs | Age, AMH, HMG)	-0.027	0.119	0.012	0.654
Corr(FSH, MII | Age, AMH, E2, HMG, Eggs)	0.001	0.949	-0.010	0.696
Corr(FSH, MII | AMH, BMI, E2, HMG, Eggs)	0.004	0.821	-0.010	0.698
Corr(FSH, MII | Age, AMH, BMI, E2, HMG, Eggs)	0.001	0.931	-0.010	0.700
Corr(FSH, MII | Age, E2, Eggs)	0.009	0.587	-0.008	0.770
Corr(FSH, MII | Age, BMI, E2, Eggs)	0.010	0.571	-0.007	0.799
Corr(BMI, HMG | Age, AMH, FSH)	0.024	0.161	-0.005	0.846
Corr(BMI, HMG | Age, AMH, FSH, Eggs)	0.025	0.137	-0.003	0.913
Corr(AMH, BMI | Age)	-0.044	0.011	0.002	0.951

4.2 Oocyte Maturation

The following pages contain supporting plots for Section II A – C in the main text.

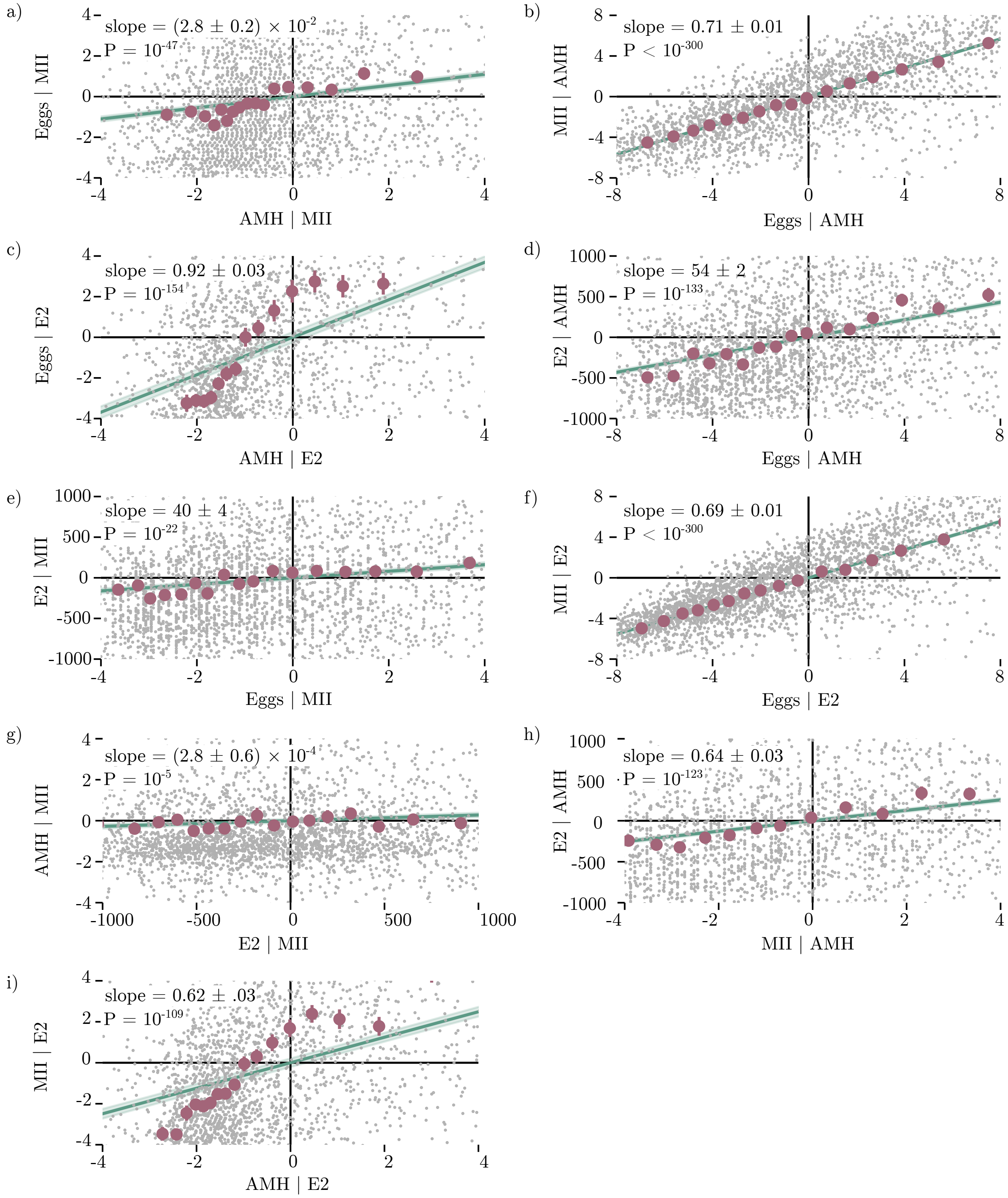


Figure 2: Additional conditional correlations corresponding to Fig. 1 in the main text. The only two conditional independencies apparent in the data are the two shown in Fig. 1 in the main text. (a) The residuals of the Eggs plotted versus the residuals of the patient AMH, after regressing both against MII, which we denote as Eggs vs Age | MII. (b) MII vs Eggs | AMH. (c) Eggs vs AMH | E2. (d) E2 vs Eggs | AMH. (e) E2 vs Eggs | MII. (e) MII vs Eggs | E2.

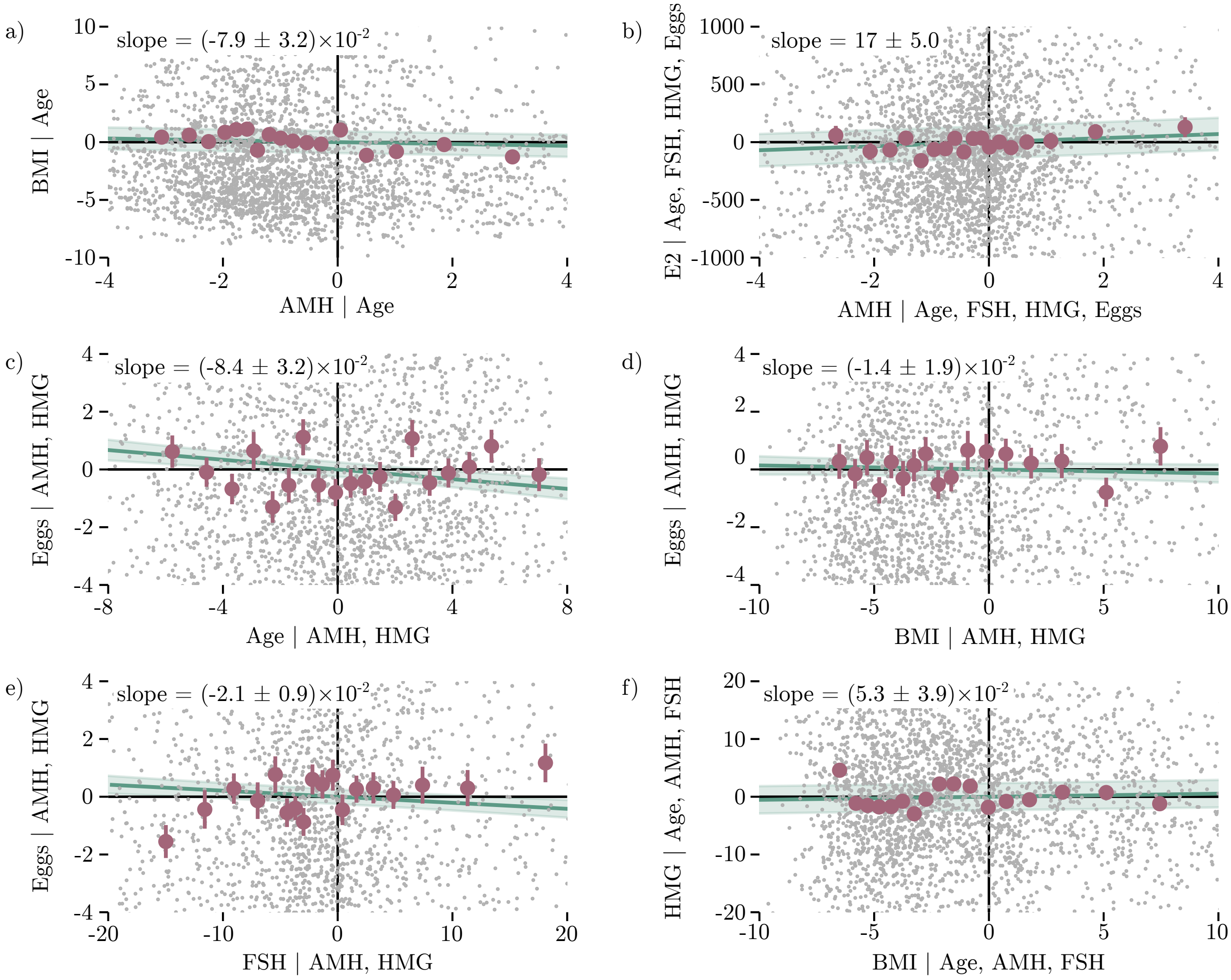


Figure 3: Conditional independencies for the graphical model shown in Fig. 3 in the main text. (a) BMI vs AMH | Age. (b) E2 vs AMH | Eggs, FSH, HMG. (c) Eggs vs AMH | AMH, HMG. (d) Eggs vs BMI | AMH, HMG. (e) Eggs vs FSH | AMH, HMG. (e) HMG vs BMI | Age, AMH, FSH.

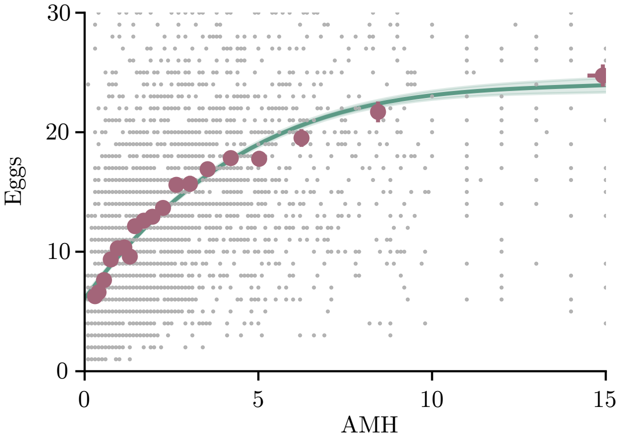


Figure 4: Eggs vs AMH. Green lines: nonlinear fit to the data; red dots: data and standard error after grouping into 20 bins; gray dots: raw data. Note that, while AMH is linear in Eggs (Fig. 3 of main text), AMH is visibly nonlinear in Eggs.

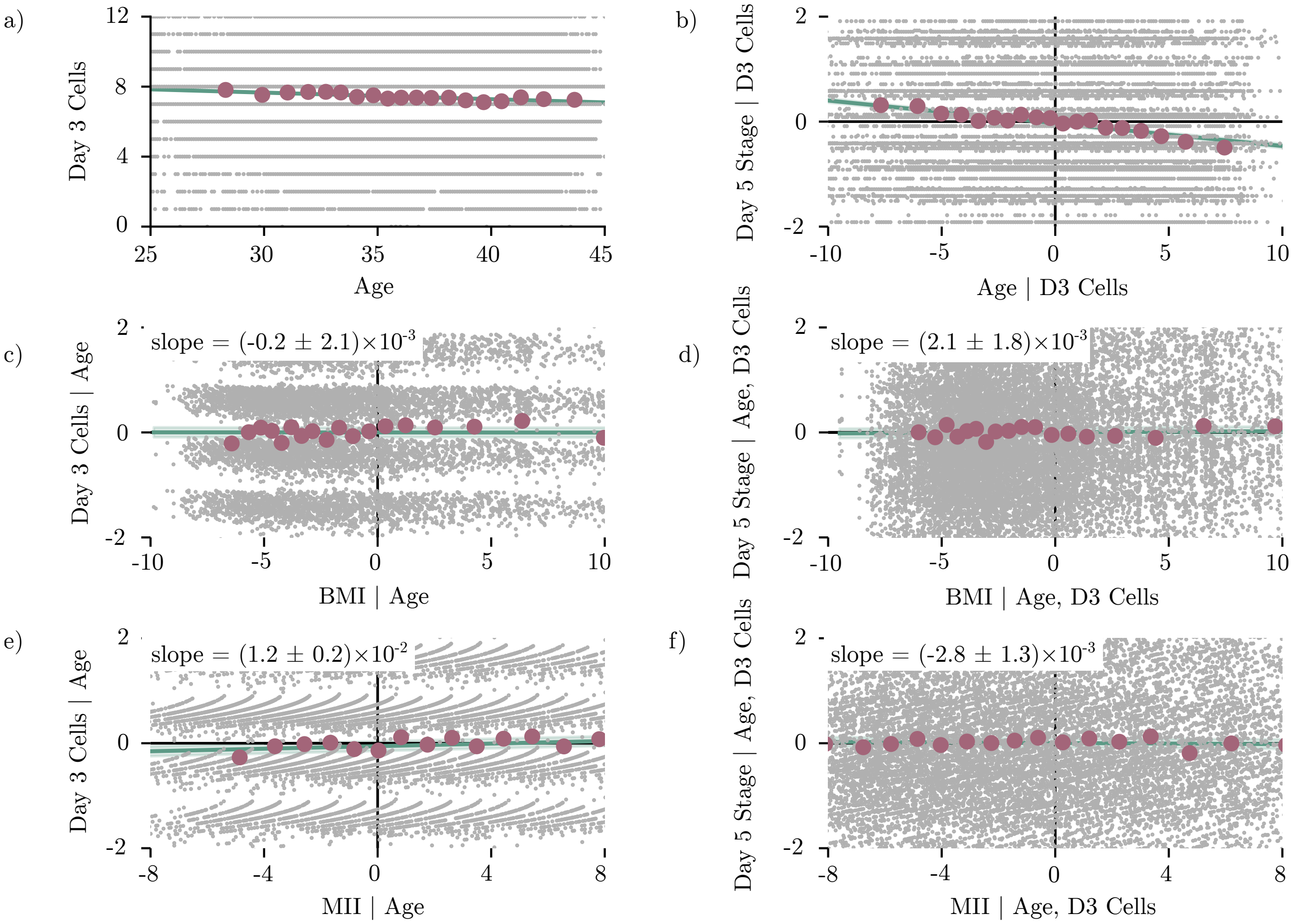


Figure 5: (a) Day 3 Cells vs Age. Green lines: linear fit to the data; red dots: data and standard error after grouping into 20 bins; gray dots: raw data. (b) Day 5 Stage vs Age, after regressing against Day 3 Cells. (c) Day 3 Cells vs BMI, after regressing against Age. The slope of the line is consistent with zero. (d) Day 5 Stage vs BMI, after regressing against Age and Day 3 Cells. The slope of the line is consistent with zero. (e) Day 3 Cells vs MII, after regressing against Age. The slope of the line is small but constrained away from zero. (f) Day 5 Stage vs MII, after regressing against Age and Day 3 Cells. The slope of the line is consistent with zero.

4.3 Embryonic Development

The data is consistent with neither BMI nor MII having any effect on the Day 5 stage of the embryo, after conditioning on Age and Day 3 Cells:

$$\begin{aligned}\text{Corr}(\text{BMI}, \text{Day 5 Stage} \mid \text{Age}, \text{Day 3 Cells}) &= 0.007 \quad (P = 0.51)\\ \text{Corr}(\text{MII}, \text{Day 5 Stage} \mid \text{Age}, \text{Day 3 Cells}) &= -0.011 \quad (P = 0.12)\end{aligned}$$

However, the data paints a slightly more complex picture for the effect of BMI and MII on Day 3 Cells. The data is consistent with no correlation between BMI and Day 3 Cells, after conditioning on Age, but suggests a very weak but nonzero correlation between Day 3 Cells and MII.

$$\begin{aligned}\text{Corr}(\text{BMI}, \text{Day 3 Cells} \mid \text{Age}, \text{Day 3 Cells}) &= -0.001 \quad (P = 0.94)\\ \text{Corr}(\text{MII}, \text{Day 3 Cells} \mid \text{Age}, \text{Day 3 Cells}) &= 0.042 \quad (P = 1 \times 10^{-5})\end{aligned}$$

While the measured correlation coefficient is nonzero between MII and Day 3 Cells (given Age), it is a tiny effect, as shown in SI Fig. 5e and as shown in Sec. 4.1, accounting for less than 0.2% of the variance in Day 3 Cells.

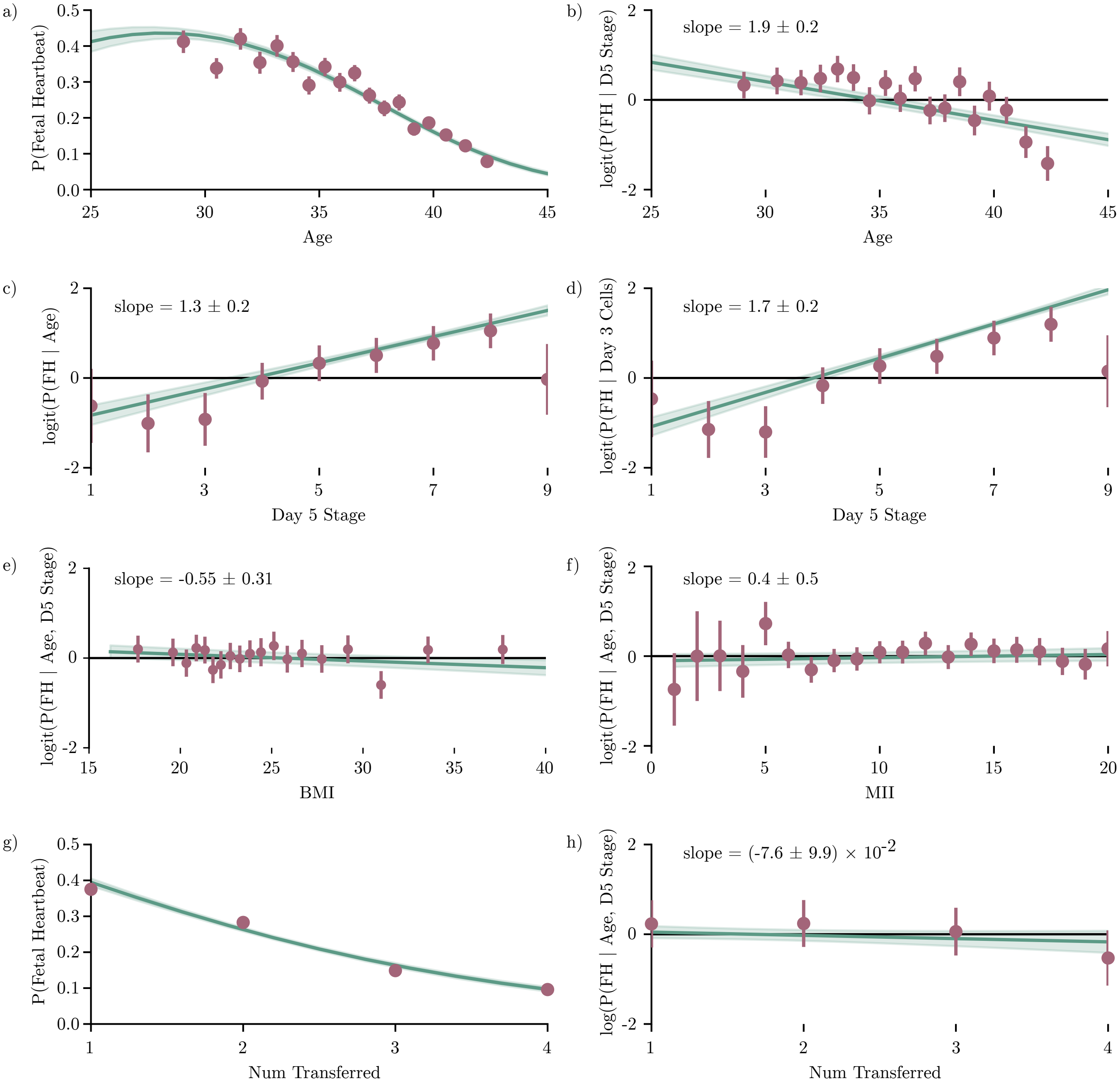


Figure 6: (a) Estimated probability of an embryo resulting in a fetal heartbeat (FH) as a function of Age alone. The red circles and errors show the probability estimated by a model that fits an independent probability of implantation for each number of cells; the green line and shaded region shows the nonlinear model with the highest model evidence and its error. (b) The logit of the estimated probability of FH as a function of Age, after regressing against Day 5 Stage. Red circles and error bars show the additional log probability estimated from a model that fits an independent logit for each value of Day 3 Cells; green line shows the best linear model for the logit and its uncertainty. Age is predictive of FH even after regressing on Day 5 Stage. (c) The logit of $P(\text{FH})$ vs Day 5 Stage, after regressing on Age. Day 5 Stage is predictive of FH even after regressing on Age. (d) The logit of $P(\text{FH})$ vs Day 5 Stage, after regressing on Day 3 Cells. Day 5 Stage is predictive of FH even after regressing on Day 3 Cells. (e) The logit of $P(\text{FH})$ vs BMI, after regressing on Age and Day 5 Stage. The data is consistent with BMI conferring no additional predictive on FH once Age and Day 5 Stage are known. (f) The logit of $P(\text{FH})$ vs MII, after regressing on Age and Day 5 Stage. The data is consistent with BMI conferring no additional predictive on FH once Age and Day 5 Stage are known. (g) Estimated probability of an embryo resulting in a fetal heartbeat as a function of the number of embryos transferred alone, and (h) the logit of $P(\text{FH})$ vs the number transferred, after regressing on Age and Day 5 Stage.

References

[1] David JC MacKay and David JC Mac Kay. *Information theory, inference and learning algorithms.* Cambridge university press, 2003.

[2] Judea Pearl. *Causality.* Cambridge university press, 2009.

[3] Gideon Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.