

# Data Aggregation and Analysis for Assessment of Ovarian Reserve

SICSA Workshop on Systems Medicine

Tom Kelsey

School of Computer Science  
University of St Andrews

[www.cs.st-andrews.ac.uk/~tom/](http://www.cs.st-andrews.ac.uk/~tom/)  
[twk@st-andrews.ac.uk](mailto:twk@st-andrews.ac.uk)



- Ovarian reserve
  - Direct and indirect indicators
  - Methods of assessment for the healthy population
- New developments
  - Systematic literature search
  - Digitized data (or data summaries)
  - Computational tools and techniques
- Recent results based on aggregated data
  - Normative & validated models of NGFs and AMH from conception to menopause
  - New hypotheses
- Implications



# Human Ovarian Reserve

- The remaining pool of non-growing follicles (NGFs) in the ovary at a given age
- NGFs are formed in large numbers in the foetal ovary with peak population occurring at 20-22 weeks gestation
  - populations decrease from this peak, largely due to apoptosis following follicular recruitment towards maturation
  - the post-menopausal ovary contains fewer than one thousand NGFs
- Human ovarian physiology is still poorly understood
  - the factors and mechanisms that initiate and control follicular recruitment and loss remain unclear
  - length of recruitment cycle?
  - location of follicles during cycle?
  - precise role of endocrine factors?
  - stem cells?



- **Direct** assessment is impossible
- No way to obtain NGF counts or rates of decline
- Longitudinal studies are impossible
- Cross-sectional studies use estimates
- Age is important
- Parity, lifestyle, genetics, ...
- **Indirect** factors are endocrine or physiological
- Anti Müllerian Hormone – AMH
- FSH
- Estradiol
- ...
- Ovarian volume
- Antral follicle counts

# New Developments

- Exponential increase in the amount of digitised data archived in searchable repositories
- Pubmed and Medline allow the identification of essentially all publications (in English) related to a given area
- Tools to collate search results, download and index digitised publications
- Older publications are being systematically digitised
- Charts, tables and figures from older publications can be semi-automatically extracted
- Once extracted, the data can be re-analysed using modern techniques
- **Datasets from multiple studies can be combined for analysis as a single dataset**

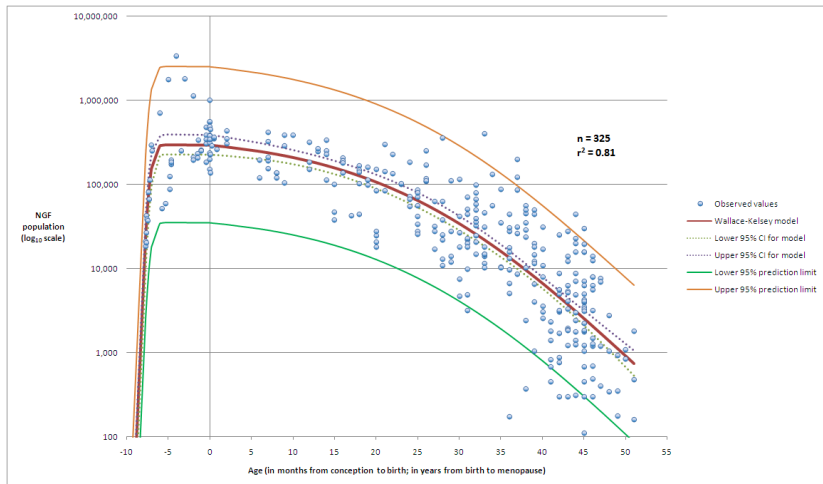


# Recent Examples

- NGF populations
  - Eight studies
  - -0.6 to 51.0 years (median 31.0)
  - $n = 325$
  - Fix zero population at conception
  - Fit plausible models
  - Validate
- Serum AMH levels
  - Twenty studies
  - -0.3 to 54.3 years (median 28.3)
  - $n = 3,260$  (from 12,241 possibles)
  - Fix zero level at conception
  - Fit plausible models
  - Validate



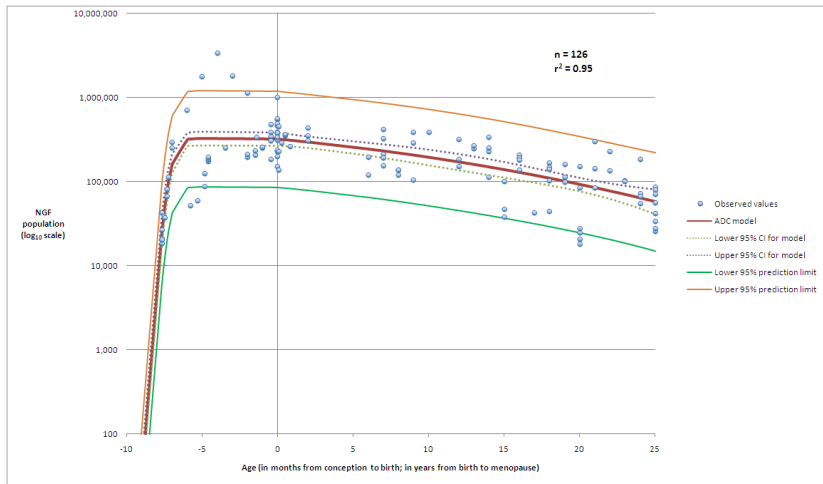
# Wallace-Kelsey Model, 2010



W H B Wallace, T W Kelsey; "Human ovarian reserve from conception to the menopause"; PLoS ONE; 5(1): e8772.



# Hypothesis 1: Age dominant factor early in life

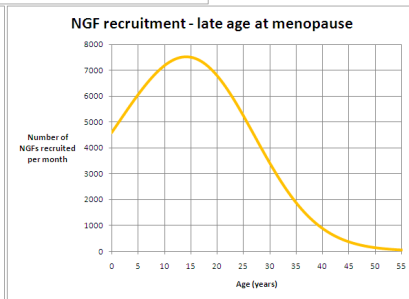
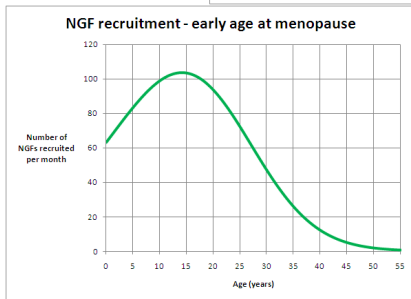
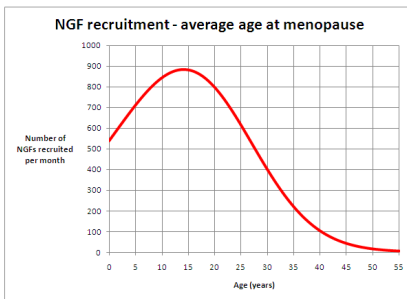


W H B Wallace, T W Kelsey; "Human ovarian reserve from conception to the menopause"; PLoS ONE; 5(1): e8772.



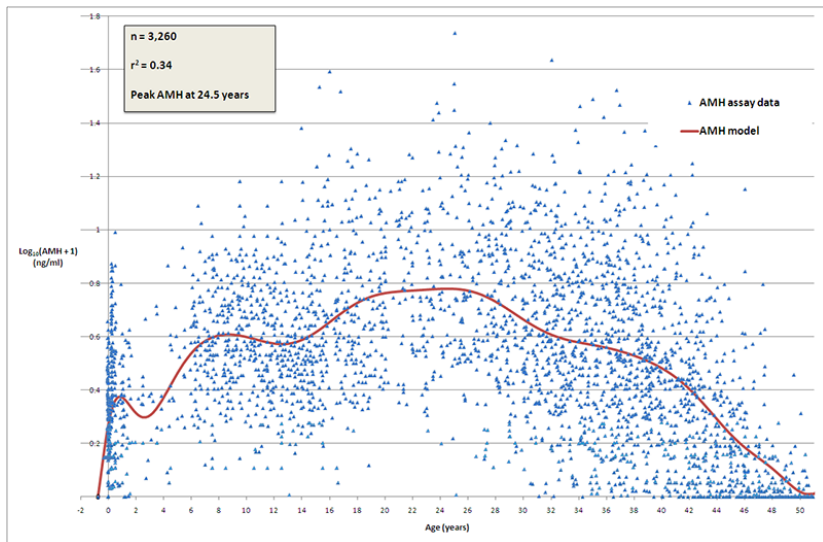


# Hypothesis 2: Activation peaks at age 13-15 years



W H B Wallace, T W Kelsey; "Human ovarian reserve from conception to the menopause"; PLoS ONE; 5(1): e8772.

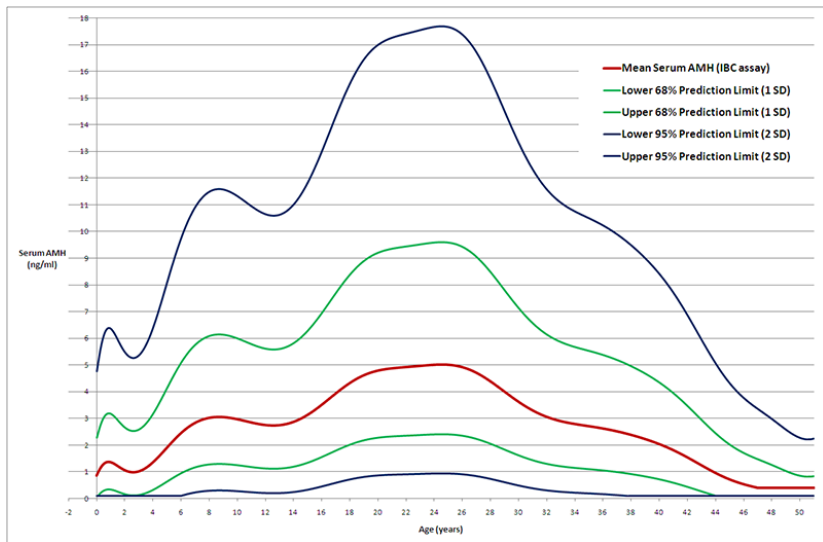
# AMH Model, 2011



T W Kelsey, P Wright, S M Nelson, R A Anderson, W H B Wallace; "A validated model of serum anti-Müllerian hormone from conception to menopause"; PLoS ONE



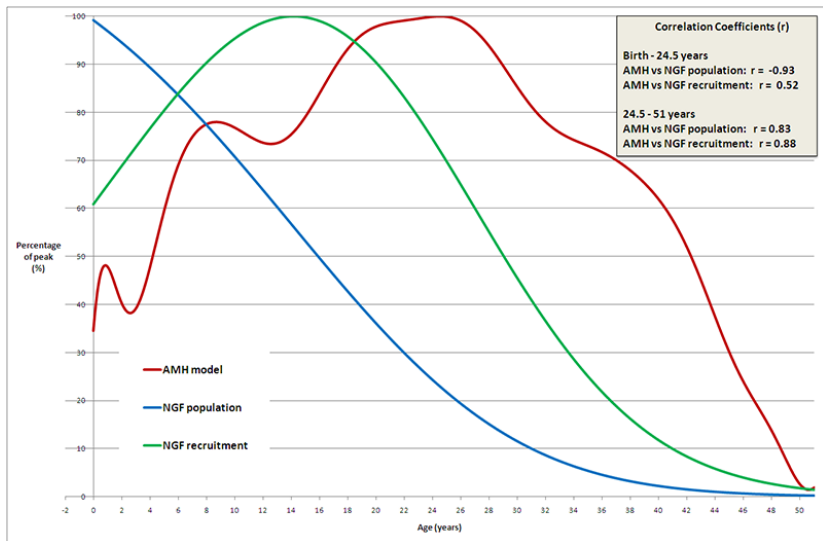
# Normative AMH for the healthy population



T W Kelsey, P Wright, S M Nelson, R A Anderson, W H B Wallace; "A validated model of serum anti-Müllerian hormone from conception to menopause"; PLoS ONE



# Hypothesis 3: AMH is a surrogate for Ovarian Reserve



T W Kelsey, W H B Wallace; "Data Aggregation and Analysis for Assessment of Ovarian Reserve"; Molecular Human Reproduction; 2011

# Implications

- "Machine Science", or data-driven approach
  - based on recent advances, but a well-known research paradigm
- Comprehensive models for the healthy population
  - confidence intervals, prediction limits & comparisons
- Hypotheses testable in future studies
  - leading to more data, improved models, insights, ...
- Inherently multidisciplinary research activity
- Data aggregation requires careful reading of the literature upon which it's based
  - Computer Science
  - Reproductive Biology
  - Reproductive Medicine
  - Mathematics



- Hamish Wallace MD & Richard Anderson MD
  - Division of Reproductive and Developmental Sciences, University of Edinburgh
- Scott Nelson MD
  - Centre for Population and Health Sciences, University of Glasgow
- Evelyn Telfer PhD
  - Department of Physiology, School of Biological Sciences, University of Edinburgh
- Luis Castillo MD, Benedicta Caserta PhD & Francisco Cóppola MD
  - School of Medicine, University of the Republic, Uruguay

