Animal experiments have contributed significantly to our understanding of mechanisms of disease.

However, their ability to predict effectiveness of treatment strategies for use in clinical trials is unclear and has remained controversial for many models.

Clinical trials are essential because animal studies do not predict with sufficient certainty what will happen in humans.
Highly cited systematic review performed to assess how often high impact animal studies translate into successful human research.

Investigators searched 7 leading journals by impact factor b/w 1980-2000 (Science, Nature, Cell) that regularly publish animal studies. Articles with > 500 citations were retrieved under the assumption that prominent findings would be tested in human trials (n=2000).

Articles were included if they investigated a preventive or therapeutic intervention in an in vivo animal model. If multiple animal studies of the same intervention were found, most cited study was retained (n=76).

For each included study, literature search identified human studies that translated the animal evidence.

“Successful” translation was defined as replication in a RCT resulting in statistically positive results according to primary outcome.

Of the highly cited animal studies, 37% were replicated in human randomized trials.

Of the interventions, 8 or one-tenth of the interventions were subsequently approved for use in patients.

Successful translation rate likely to be significantly lower.

Multiple reasons for the failure to translate findings from animal models to clinical trials.

Internal validity refers to the extent to which design and conduct of the trial eliminate possibility of bias.

In clinical trials, several systematic reviews and meta-analyses have demonstrated that inadequate methodological approaches in controlled clinical trials are associated with bias.

However, in animal studies, the impact has been examined less extensively.
In a metaanalysis of animal studies investigating interventions in emergency medicine, odds of positive result were more than three times as large if the publication did not report randomization or blinding compared with publications that did.1

In systematic review of animal models of hypothermia for acute ischemic stroke, an inverse relation was found between study quality as assessed by a study-quality checklist and effect size.2

External validity refers to the extent to which the results of an animal experiment provide a correct basis for generalizations to the human condition.

- Reductions in external validity:
  - models with diseases or injury that are dissimilar to the human condition;
  - induction of disease under study in animals that are young and healthy while patients are elderly and with comorbidities;
  - homogenous group of animals versus heterogeneous group of patients;
  - use of male or female animals only;
  - treatment time course or doses that are unrealistic in humans;
  - differences in outcome measures and timing of outcome assessment.

The Inflammation and Host Response To Injury, Large Scale Collaborative Research Program was funded to study early inflammatory response to serious injuries.

Investigators performed genome-wide expression analysis on white blood cells from hundreds of burn, trauma, or sepsis patients and found patterns of gene expression that predicted outcome.

However, attempts at publishing these findings resulted in a common objection that the researchers had not shown the same gene response had happened in mice.

"They were so used to doing mouse studies that they thought that was how you validate things. They are so ingrained in trying to cure mice that they forget we are trying to cure humans."  

"That started us thinking," he continued. "Is it the same in the mouse or not?"

Genomic responses in mouse models poorly mimic human inflammatory diseases

- Murine models have been extensively used in recent decades to identify drug candidates for subsequent human trials but few have shown success.
- Authors discuss that of 150 clinical trials evaluating agents intended to block inflammatory response in critically ill patients, every single one has failed.
- Goal was to compare the genomic response seen in human inflammatory diseases and a commonly used murine model.
  - Prior to the paper, no studies had evaluated on a molecular basis how well the murine clinical models mimic human inflammatory diseases in patients.

PATIENT ENROLLMENT AND SAMPLING

- Between 2003-2009, 167 patients admitted to US Level I Trauma Center (1 of 7) with blunt injury associated with SBP < 90, or HCO3 deficit > 6 mEq/L, transfusion requirement < 12 h, Abbreviated injury score > 2 had blood samples taken within 12 h and 1, 4, 7, 14, and 21 d after injury.
- Between 2000-2009, 244 burn patients were enrolled if admitted within 96 h of injury, TBSA > 20%, and required at least 1 excision and grafting procedure and had blood draws up to 1 year after injury.

PATIENT ENROLLMENT AND SAMPLING

- Eight healthy male and female subjects between 18-40 years of age received either IV endotoxin E. coli 0113 or 0.9% NS and blood samples were collected at 0 hours, 2, 4, 6, 9, and 24 hours after infusion.
- In addition, 35 healthy control subjects were recruited between 2004 and 2007.

MALE MURINE INJURY MODELS

- All mice were male C57BL/6J.
- Groups of 12 mice undergoing injury were giving anesthesia with isofluorane and then subjected to 25% TBSA scald burn or trauma/hemorrhage. T/H consisted of laparotomy followed by withdrawal of sufficient blood from arterial line to decrease and maintain MAP at 45 mm Hg for 90 minutes.
- After burn injury, mice were resuscitated with 1 mL NS IP. After T/H, mice were resuscitated with Ringer’s at 4x shed blood volume.
- Mice receiving LPS (10 ng E. coli 0113) were not anesthetized.
METHODS

- Total blood leukocytes were isolated and total cellular RNA was extracted and hybridized onto an Affymetrix GeneChip.
- Baseline represented mean gene expression controls. Maximum fold change (fold change) refers to maximum deviation from baseline.
- Response time was defined as time to one-half maximum fold change. Recovery time was defined as time from baseline to one-half of maximum.

RESULTS

- 5,544 genes were identified as significant between patients and healthy subjects and 4,918 of these had mouse orthologs.
- Among the 4,918 genes, correlations of maximum gene changes (Pearson correlation, $R^2$) and percentages of genes changed in the same direction in all six conditions were compared.
  - The Pearson correlation is +1 for direct linear relationships, -1 for inverse relationships, and 0 for no relationship.
  - By random chance, 50% of genes between two uncorrelated conditions are expected to change in the same direction.
- High correlation in gene response between human trauma and burns ($0.91, 97\%$) and moderate similarity between human endotoxemia and burn/trauma ($0.47, 88\%$).
- Low correlation in gene response between all three murine models of injury ($R<0.13, 39-63\%$).
- Low correlation in gene response between all human injuries and mouse models ($R<0.09, 47-63\%$).
RESULTS

Next, investigators sought to evaluate differences in both response and recovery time of gene expression between the models.

In all conditions, gene response time occurred within the first 6-12 hours.

However, recovery times were significantly different.
- In murine models, recovery occurred within hours to 4 days;
- In humans, recovery time was 1-6 months.
RESULTS

Major signaling pathways regulated in human injuries were identified and compared across models with human burns as a reference.

- Innate immunity upregulated and adaptive immunity downregulated.

Evaluation of the top 40 pathways demonstrated median $R^2$ of 0-0.16 and gene percentages of genes changed in the same direction of 52-67%.

- Investigators evaluated additional patient and corresponding mouse model studies from Gene Expression Omnibus (GEO) for other acute inflammatory diseases.

- Maximum fold changes from the data sets were compared with the 5,554 genes significantly changed in the three human injuries to calculate correlations and directionality of gene response.

RESULTS

Genomic responses in humans correlated well with each other with $R^2$ from 0.47-0.91 and % of genes changing the same direction from 83-97%.

In mice, $R^2$ ranged from 0-0.08 with 47-61% of genes changing in the same direction.

RESULTS

"Studying disease in patients is much more complex than studying model systems."

- This may be true but it may be that model systems and human disease are equally complex. The issue may be that model systems and disease are just dissimilar.

- Two broad points of the article.

DISCUSSION
Critically ill patients who suffer different injuries often have similar appearing physiologic reactions, a condition known as Systemic Inflammatory Response Syndrome (SIRS).

An concept central to drug testing for this syndrome has been that the molecular mechanisms underlying it are similar regardless of the etiology.
- High correlation in response between trauma, burns, and endotoxemia in humans despite heterogeneity tends to support this hypothesis.

Genomic response of peripheral WBC in the C57BL/6J mouse strain poorly reflects human inflammatory diseases.
- Why might this be the case?

The authors stress the point that the evolution of the immune system for any species is a consequence of microbe-exerted selection pressure for that species.
- In endotoxin model, mice are highly resilient where lethal dose is 5-25 mg/kg for most strains of mice, whereas a dose 1,000,000-fold less (30 ng/kg) causes shock in humans.

This difference in sensitivity may result in genomic responses that reach an upper threshold in human disease, whereas in mice it doesn’t reach threshold and results in greater heterogeneity in response.
- It is possible that a higher level of injury in mouse models would result in a transcription response that actually mimics the response seen in patients.
DISCUSSION

- Multiple other potential reasons that the C57BL/6J model poorly mimics human injury
  - evolutionary difference between mice and humans,
  - complexity of human disease,
  - inbred nature of the mouse model,
  - differences in cellular composition between mouse and human tissues can contribute to differences seen in the molecular response;

- While the reasons for the differences are potentially interesting to study, ultimately this paper suggests that C57BL/6J is a poor strain for drug development for these inflammatory conditions.

- This is not irrelevant because C57BL/6J is often used to study inflammatory conditions.

DISCUSSION

- Multiple other potential reasons that the C57BL/6J model poorly mimics human injury
  - similarities in cellular composition between mouse and human tissues are not reflected in peripheral WBC genomic expression patterns,
  - different temporal spans of recovery from disease,
  - late events related to clinical care of patients (fluids, drugs, surgery) may alter genomic responses not captured in murine models,
  - differences in gender,
  - isofluorane may have suppressed the immune response\(^1\).


DISCUSSION

- How to proceed?
  - The authors note that a practical approach forward would be to require greater validation of animal models.
  - Studying whether a model mimics or fails to mimic the molecular or cellular behavior of key genes and pathways relevant to human disease would be critical to advance models for drug development.
DISCUSSION

- New approaches
  - outbred mice with greater genetic heterogeneity;
  - comprehensive genomic descriptions in patient studies to define human disease which could then help guide development of animal models;
  - may require use of other organisms that model human disease more accurately, although still require validation;
  - tissue chips – miniature 3D organs made with living human cells
  - NIH is dedicating $70 million over 5 years
  - "humanized" mice.

"As a cornerstone of modern biomedical research, the use of mouse models has dominated scientific literature. The prevailing assumption—that molecular results from current mouse models developed to mimic human diseases translate directly to human conditions—is challenged by this study."

REACTION

- "For decades, mice have been the species of choice in the study of human diseases. But now, researchers report evidence that the mouse model has been totally misleading for at least three major killers—sepsis, burns and trauma. As a result, years and billions of dollars have been wasted following false leads, they say."

- "The good news for mice is that humans have spent billions of dollars to solve their illnesses. But it seems researchers have tortured mice in vain for decades in the search for drugs to help humans recover from certain traumas..."

- "The study's investigators tried for more than a year to publish their paper. They submitted it to the publications Science and Nature...reviewers did not point out scientific errors but "the most common response was, 'It has to be wrong. I don't know why it is wrong, but it has to be wrong.'"
If these results are validated, how should they change practice?
- PETA
- Jackson Labs

‘Of mice and not men’ not a novel concept, yet this article has promoted a great deal of discussion.
- Is it because the use of this technique is new and more compelling than previous comparisons?
- Is it because the narrative the authors constructed – investigators losing sight of human disease, bias in the peer review process – is a compelling one?

REFERENCES
2. Hackam and Redelmeier. Translation of research evidence from animals to humans. JAMA 2006; 296: 1731-1732.

QUESTIONS?