# Heather Battey's invited discussion of 'Regression by composition' by Farewell et al.

Heather Battey*

### Foundations of modelling

In McCullagh's (1999, 2002, 2022) formalisation of the foundations of modelling, a treatment has an effect on the probability distribution of the outcome. The untreated outcome distribution belongs to a family $\{P_\gamma : \gamma \in \Gamma\}$ and the treatment induces a transformation of this set via a group $\mathcal{G}$ which, in the simplest models, acts directly on $\Gamma$. Allowing $\gamma_i$ for unit $i$ to depend on intrinsic features $w_i$ recovers standard regression formulations, e.g. $\gamma_i = w_i^{\mathrm{T}}\beta$ for $\Gamma = \mathbb{R}$, and $\gamma_i = \exp(w_i^{\mathrm{T}}\beta)$ for $\Gamma = \mathbb{R}^+$, where the group actions on $\Gamma$ are addition and multiplication respectively. While the group element defining the treatment effect is stable over units, the control distribution is unit-specific in view of $\gamma_i$. There are elaborations for interaction.

Farewell et al. extend this group theoretic formalisation, treating the effect of all variables as group elements, with a possibly different group action for each effect. If different group actions are used for different variables, then an ordering is inherent. This makes the formulation somewhat unnatural for intrinsic variables fixed at baseline, but might have some appeal for successive treatment/exposure variables and intermediate outcomes.

### Equivalence classes and confidence sets of models

By analogy with Gaussian graphical models, where multiple causal explanations (Markov equivalence classes) give rise to the same distribution over the outcome, one might expect similar considerations to arise here. In other words, in models with intrinsic variables treated symmetrically and the effects of multiple treatments or intermediate outcomes modelled via the compositional approach, can there be multiple causal models, a priori plausible, that are statistically indistinguishable? How should model adequacy be assessed?

### Instability induced by maximum-likelihood fitting

A drawback of most collapsible models is that they are not physical, e.g. they can produce probabilities that fall outside of the $[0, 1]$ range. In other words, the model is a poor approximation for some individuals, and whether it constitutes a model is up for debate. Enforcing physical constraints leads to a parameter space that depends on the sample chosen, undermining the stable interpretation of parameters and violating the McCullagh (2002) axioms. It seems important, therefore, to allow fitted values to fall out of range for some observations, which is not permissible with maximum-likelihood fitting. Do such considerations play a role in regression by composition?

### Interpretation in collapsible models

Regression-type formulations interpreted purely mathematically via partial differentiation ("all other variables notionally held fixed") invite dangers of causal claims, the security of which

*Department of Mathematics, Imperial College London, UK, h.battey@imperial.ac.uk.

depends strongly on context. Interpretation is most delicate for coefficients on intrinsic variables. Interpreted mathematically, linear in probability models have regression coefficients quantifying the expected change in the proportion of $m$ individuals likely to experience the positive outcome in response to a hypothetical replacement of $m$ individuals who differ by one unit in the $j$th component and are otherwise identical, $m$ being arbitrary. This switches the perspective from individual-level quantities to population-level proportions: variables that were intrinsic at the individual level are not so when treated as population averages. I would be interested to know the authors' views. What manifestations, if any, does this have in the collapsible components of RBC?

The collapsible, but physically implausible, linear in probability model appears to be the only model for binary data for which the model-based and Neyman-Rubin model-free definitions of treatment effect agree under the fictitious idealisation discussed by Battey and Edgar (2025+). For any other generating process, the Neyman-Rubin treatment effect is inherently unstable with respect to the sampling of new individuals, and its relevance to scientific understanding seems questionable (see also Cox, 1958a, 2012). This, however, is not in my view an argument for collapsible models, but rather an argument against the Neyman-Rubin treatment effect.

### Relevance of collapsibility

Discussions of collapsibility hinge on an omitted baseline covariate being treated as a random variable, which, in the context of regression, is at odds with ancillarity considerations.

It seems to me that if covariates thought to be relevant are measured, then they ought to be included. Does this not make the argument for collapsibility rather weak? A case for deliberately omitting them would be if they were too numerous, pointing back to assessments of model adequacy.

### Collapsibility and unmeasured covariates

Suppose that relevant baseline covariates are not measured. In the special case of matched pairs of individuals, the baseline probabilities within pair are equal. Cox (1958b) introduced, in the context of a logistic model, an all-encompassing nuisance parameter $\gamma_i$ specific to the $i$th pair, giving rise to a special case of McCullagh's formulation

$$\begin{aligned} q_{1|0}^{(i)} := \text{pr}(\text{success} \mid \text{control}) &= \frac{e^{\gamma_i}}{1+e^{\gamma_i}}, \qquad q_{0|0}^{(i)} = 1 - q_{1|0}^{(i)} \\ q_{1|1}^{(i)} := \text{pr}(\text{success} \mid \text{treated}) &= \frac{e^{\gamma_i+\beta}}{1+e^{\gamma_i+\beta}}, \quad q_{0|1}^{(i)} = 1 - q_{1|1}^{(i)}. \end{aligned}$$

By appealing twice to ancillarity considerations, first in the $n$ pairwise tables, then in the discordant-pair table, all $n$ nuisance parameters $(\gamma_i)_{i=1}^{n}$ are eliminated and reliable inference on $\beta$ can be performed using a suitable conditional distribution. Thus, although any collapsed $2 \times 2$ table leads to erroneous inference on $\beta$, it seems possible in certain circumstances to evade such issues through a careful analysis.

A little-known reference concerning collapsibility is Cox (2007) which expounds the apparent dependencies induced through marginalisation in quantile, proportional hazards, and discrete regression relationships. It is intermediate outcome variables, however, that are treated as random, not baseline covariates.

It is a pleasure to thank the authors for an original, scholarly and thought-provoking contribution to an important topic.

## REFERENCES


Battey, H. S. and Edgar, C. (2025+). Treatment effect: a critique. *arXiv: 2601.15467.*

Cox, D. R. (1958a). The interpretation of the effects of non-additivity in the Latin square. *Biometrika*, 45, 69–73.

Cox, D. R. (1958b). Two further applications of a model for binary regression. *Biometrika*, 45, 562–565.

Cox, D. R. (2007). On a generalization of a result of W. G. Cochran. *Biometrika*, 94, 755–764.

Cox, D. R. (2012). Statistical causality: some historical remarks. Pages 1-5 in *Causality: statistical perspectives and applications.* Edited by C. Berzuini, P. Dawid, L. Bernardinelli. Wiley, Chichester.

McCullagh, P. (1999). The algebraic structure of generalised linear models. University of Chicago, technical report number 489.

McCullagh, P. (2002). What is a statistical model? *Ann. Statist.*, 30, 1225–1310.

McCullagh, P. (2022). *Ten Projects in Applied Statistics.* Springer, Cham.