

Conclusie over optimale schaal voor Nederlandse ziekenhuizen kan nog steeds niet getrokken worden.

Reactie op Naschrift: “Nederlandse ziekenhuizen toch echt te groot”

Marc Pomp, Zelfstandig adviseur gezondheidseconomie (info@marcpomp.nl)

Jan-Peter Heida, partner bij SiRM (jp.heida@sirm.nl)

10 december 2017

In mei 2017 publiceerden Bart van Hulst en Jos Blank (hierna HB) het artikel “Nederlandse ziekenhuizen te groot voor verdere schaalvoordelen”. Hierop reageerden wij met de reactie “Conclusie over optimale schaal kan niet getrokken worden”. HB hebben daar een naschrift bij geschreven, getiteld “Nederlandse ziekenhuizen toch echt te groot”. Onze reactie en het naschrift worden beiden gepubliceerd in de ESB van 21 december 2017 (online gepubliceerd op 6 december 2017).

HB concluderen dat Nederlandse ziekenhuizen te groot zijn voor verdere schaalvoordelen. Op basis van een regressieanalyse van studies naar schaalvoordelen stellen zij dat 320 bedden een optimale schaal is. Aangezien de meeste Nederlandse ziekenhuizen groter zijn, zouden meer ziekenhuisfusies geen schaalvoordelen opleveren.

In onze reactie op hun artikel stellen wij dat de conclusie niet getrokken kan worden. De gemiddelde omvang van de ziekenhuizen in de achterliggende studies is kleiner dan het gemiddelde Nederlandse ziekenhuis. Bovendien is niet afdoende gecorrigeerd voor patiëntkenmerken, en is er geen rekening gehouden met kwaliteit van zorg. Niet parametrische studies wijzen bovendien op onbenutte schaalvoordelen bij grote ziekenhuizen. En tot slot kan het gebruik in een meta-analyse van de gemiddelde ziekenhuisomvang resulteren in een onderschatting van de optimale schaal.

In hun naschrift op onze reactie stellen HB dat onze redenering niet correct is.

Wij zetten hieronder uiteen op welke punten we het niet eens zijn met de op- en aanmerkingen die HB aangaande onze reactie maken in hun naschrift. We behandelen de punten in de volgorde van het naschrift. Citaten uit hun naschrift staan hieronder in vet en tussen aanhalingstekens.

Gaat de daling door boven 450 bedden?

“PH doen de suggestie dat na een daling en een stijging van de gemiddelde kosten (U-curve) ze vervolgens misschien weer dalen”

Deze opmerking kunnen wij niet plaatsen. Onze figuur 1 heeft als x-as het aantal bedden en als y-as de schaalelasticiteit. Het betreft dus niet de gemiddelde kosten bij een bepaald aantal bedden (waarvoor in de literatuur inderdaad de U- en L-curve voorkomen), maar de afgeleide van de gemiddelde kosten bij dat aantal bedden. Integratie van die curve levert een U-vormige curve, vandaar onze opmerking in de reactie: “Dit is precies het patroon dat volgens figuur 1 in HB uit de studies naar voren komt”¹. (Zie ook verder onder de kop “De schaalelasticiteiten gelden niet bij het gemiddelde”).

“Tot slot laat figuur 1 van ons artikel zien dat boven de 270 bedden geen resultaten worden gevonden met schaalvoordelen. Er is dus geen studie te vinden die de stelling van PH ondersteunt.”²

Wij zijn van mening dat figuur 1 van HB onze stelling juist illustreert. We stellen namelijk niet dat er per se schaalvoordelen boven 450 bedden zouden zijn. We stellen dat je uit het gepresenteerde onderzoek niet kunt concluderen dat die schaalvoordelen er al dan niet zijn. Bovendien stellen we dat eventuele conclusies uit het onderzoek niet zijn toe te passen op de Nederlandse situatie. Dat blijkt juist heel mooi uit figuur 1. Er zijn namelijk maar 2 van de 41 studies die in de buurt komen van de gemiddelde schaal van Nederlandse ziekenhuizen (in 2016 gemiddelde: 466 bedden; mediaan: 379³).

Geen goede correctie patiëntkenmerken

“PH kiezen voor de gewogen-outputindicator en stellen dat je alle andere studies eigenlijk moet weggooien”.

Wij zijn ons er niet van bewust dat we suggereren studies weg te gooien. We stellen voor om ze anders te gebruiken. We merken op dat het effect van casemix voor deel terecht kan komen in de coëfficiënt van het aantal bedden. Die coëfficiënt is dan niet alleen een maat voor de impact van schaal, maar ook van casemix. Dat is wellicht nader te onderzoeken door een interactieterm op te nemen met de wijze waarop voor casemixverschillen is gecorrigeerd.

¹ Schaalelasticiteit boven 1 bij laag aantal bedden, dus gemiddelde kosten dalen met het aantal bedden, vervolgens een gebied met schaalelasticiteit gelijk aan 1 (dus richtingscoëfficiënt 0), de bodem van de U-curve en vervolgens schaalelasticiteit lager dan 1 met dus een toename van de gemiddelde kosten bij toename van het aantal bedden.

² In het naschrift van Hulst en Blank staat dit onder het kopje ‘Geen goed correctie patiëntkenmerken’.

³ www.jaarverantwoordingzorg.nl Dataset verantwoordingsgegevens zorginstellingen 2016.

“De keuze van PH is echter gebaseerd op een foutieve interpretatie van significantie bij het gebruik van dummyvariabelen met een referentiecategorie.”

Wij begrijpen deze opmerking niet. Wij gaan er van uit dat de coëfficiënt horend bij een bepaalde dummyvariabele altijd een effect ten opzichte van een referentiecategorie weergeeft. En die coëfficiënt is al dan niet significant.

“Ook suggereren PH dat een nog betere casemixcorrectie bestaat met nog hogere schaalearlasticiteit.”

Mochten we die suggestie hebben gewekt, dan is dat spijtig. Dat was zeker niet onze bedoeling. Sterker nog, wij waarschuwen er juist voor dat het heel moeilijk is om casemix correcties goed door te voeren. Het commentaar van gezondheidseconoom Newhouse uit 1994 is hierop nog steeds van toepassing⁴: “Case-mix controls, usually Diagnosis Related Groups (DRGs) or something more aggregated such as medical-surgical cases, are problematic. The likelihood that variation within a DRG is not random by hospital is well known.”.

Kwaliteitsverschillen niet meegenomen

“PH veronderstellen betere kwaliteit in grote ziekenhuizen, maar onderschrijven dat dit niet aantoonbaar is.”

We veronderstellen dit niet. We sluiten het alleen niet uit. Het is wel degelijk mogelijk dat in grotere ziekenhuizen met hogere volumes van behandelingen de kans op hogere kwaliteit hoger is dan in kleinere ziekenhuizen die minder van die behandeling uitvoeren.

Bovendien halen veel kleinere ziekenhuizen de volumenorm voor bepaalde behandelingen niet. Zij voeren die behandelingen dan ook niet uit. Dit geldt overigens niet alleen voor behandelingen met minimumnormen; grote ziekenhuizen leveren behandelingen die kleine ziekenhuizen niet leveren. HB doen dus uitspraken over schaalears effecten van organisaties die verschillende producten leveren.

Niet-parametrische studies leiden tot andere conclusie

“De opmerking van PH dat niet-parametrische studies vaak de optimale schaal rapporteren, is feitelijk onjuist.”

Terecht punt. We hebben eroverheen gelezen dat VH deze optimale schaal zelf heeft berekend door aanvullende aannames te maken. Aannames die hij zelf overigens onrealistisch vindt:

“We first assumed that IRS, CRS and DRS hospitals are neatly ordered, such that IRS hospitals are smaller than CRS hospitals and CRS hospitals are smaller than DRS hospitals. Note that in practice this will seldom be true: there will normally be an overlap between the range of IRS and CRS, CRS and DRS, and possibly also IRS and DRS hospitals. Second, we assumed that in each study hospital sizes are normal distributed.” (p. 83).

⁴ Newhouse, Joseph P, “Frontier estimation: How useful a tool for health economics”, J. of Health Economics 13 (1994) p319

“Omwille van de volledigheid is een analyse gemaakt door, met pijn een moeite, voor nog der-tien studies de optimale schaal te imputeren. Op basis hiervan wordt de optimale schaal op 220 bedden geschat. Vreemd genoeg melden PH deze uitkomst niet. Indien niet-parametrische studies al tot een andere conclusie leiden, dan is dit dat de optimale schaal nog kleiner is.”

Dit kunnen wij niet plaatsen. VH schrijft in zijn proefschrift “There is a wide variation for the optimum scale: it varies from 22 beds to a range that stretches to 430 beds”. Het gemiddelde is inderdaad 220, maar dat gemiddelde is niet “de” optimale schaal. Dat is nu juist de reden waarom wij de resultaten rapporteren van de regressie-analyse van Van Hulst, waarin hij probeert te verklaren waarom de verschillende niet-parametrische studies tot verschillen in optimale schaal komen. Hij concludeert:

“The only parameter that is significant at the 1% level, is the parameter estimate for the average number of beds. There is a strong correlation between the average number of beds and the optimum. The regression results indicate that for each additional bed, the optimum scale increases by 0.72 beds. In practice, things are a bit more nuanced. The average number of beds also captures differences in context. The results indicate that an optimum is sensitive to the context. Therefore, we should keep in mind that for specific hospital types, the optimum scale might very well differ from the general optimum as calculated previously.” (p. 86).

Ofwel: bij de non-parametrische studies neemt de optimale schaal juist toe met het gemiddeld aantal bedden. Maar deze uitkomst wordt door vH terzijde geschoven omdat verschillen in het gemiddeld aantal bedden ook verschillen in context zouden oppikken, waarmee kennelijk bedoeld wordt op verschillen in type ziekenhuis en verschillen in casemix. Dit is zeker plausibel, maar dat geldt dan toch net zo goed voor de beddenvariabele in de parametrische studies die de basis vormen voor het ESB-artikel?

“De methode van Wilson en Carey leent zich slecht voor het afleiden van schaaffecten”

Waarom? Wilson en Carey (2004)⁵ concluderen op grond van een niet-parametrische onderzoek dat ook bij de grootste ziekenhuizen in hun data (meer dan 400 bedden) verdere groei schaalvoor-delen kan opleveren:

“Our results differ from those in the existing literature on hospital costs. We find evidence of increasing returns to scale among hospitals above the median size, extending to the largest decile in terms of size, in our analysis of EPSE. The past decade has seen a large number of hospital mergers, including mergers between large hospitals. Our results suggest that these hospitals are exploiting economies of scale.” (p. 520)

⁵ Wilson, P.W., Carey, K. “Nonparametric analysis of returns to scale in the us hospital industry”. *Journal of Applied Econometrics*, 19(4) (2004),

Van Hulst noemt de studie van Wilson en Carey in zijn proefschrift, en ook hun fundamentele kritiek op de parametrische aanpak:

“Wilson and Carey (2004) deal explicitly with scale effects. Their conclusion is that, due to possible misspecifications, studies that apply the parametric method conclude too early that diseconomies of scale may occur.” (vH p. 37).

Van Hulst geeft echter niet aan of hij het met deze kritiek eens of oneens is. Dat is problematisch, want door fundamentele kritiek op je methode te noemen zonder die te weerleggen, blijft de lezer in het ongewisse of je de kritiek wel kunt weerleggen.

De schaalelasticiteiten gelden niet bij het gemiddelde

“De schaalelasticiteiten gelden niet bij het gemiddelde: Meestal gaat het dan om de gemiddelde schaal in die steekproef (de schaalelasticiteit wordt geëvalueerd in het gemiddelde).”

Wat hier tussen haakjes staat is precies ons punt. De schaalelasticiteit wordt berekend op basis van een regressie. Als het onderliggende verband niet lineair is maar bijvoorbeeld zoals in onze figuur 1, dan geldt de regressie-coëfficiënt niet voor het gemiddelde aantal bedden in de steekproef. Dit geldt overigens ook voor een andere vorm van deze curve zolang die niet symmetrisch is, of, mocht die toch symmetrisch zijn, als de steekproef niet representatief is voor de gehele curve. De kans is zeer klein dat aan beide voorwaarden is voldaan.

Conclusie

“Uit de meta-analyse blijkt dat verschillende methodes, contexten, specificaties en gegevens weliswaar van invloed zijn op de resultaten, maar uiteindelijk is er een verrassende consensus over schaalears effecten.”

Op p.79 in het proefschrift schrijft vH veel minder stellig:

“For a reference study, based on the most common characteristics, the optimum size is 320 beds. However, it is hard to really pinpoint the optimum scale since the optimum scale depends on several factors, including study characteristics.” (cursivering toegevoegd).

Ook andere onderzoekers zijn heel wat minder stellig:

- Wilson en Carey (2004): “Given the volume of hospital mergers and acquisitions, there is remarkably little consensus on the extent of scale economies in hospitals.” (p 506, cursivering toegevoegd).
- Giacotti et al. (2017)⁶ publiceerden vrijwel gelijktijdig met HB een literatuuroverzicht in het wetenschappelijke tijdschrift PLOS one, met deels hetzelfde doel als HB. Deze studie komt tot de volgende conclusie: “Studies analysed in this review showed that economies of scale

⁶ Giacotti, Monica, Annamaria Guglielmo, and Marianna Mauro. "Efficiency and optimal size of hospitals: Results of a systematic search." PLoS one 12.3 (2017).

are present for merging hospitals. Results supported the current policy of expanding larger hospitals and restructuring/closing smaller hospitals. In terms of beds, studies reported consistent evidence of economies of scale for hospitals with 200±300 beds. Diseconomies of scale can be expected to occur below 200 beds and above 600 beds.”

Wij hopen dat we onze reactie op het artikel van van Hulst en Blank (“Nederlandse ziekenhuizen te groot voor verdere schaalvoordelen”) verder hebben kunnen verduidelijken door in te gaan op het naschrift.

Volgens ons kan op basis van het in het artikel gepresenteerde onderzoek niet worden geconcludeerd dat Nederlandse ziekenhuizen te groot zijn.